

Divisions in Euroland

Twelve European nations this month adopted a common currency. But the goal of creating a unified scientific career structure remains distant. The consequent brain drain will undermine Europe's ability to compete with the United States.

The euro represents a political leap of faith that Europe's leaders hope will boost the continent's economic power. Even if it succeeds in that regard, however, the single currency may do little to stop the brain drain of young scientists from the continent. The three largest nations within 'Euroland' — Germany, France and Italy — each boast a rich scientific heritage. But many of their young scientific high-flyers cast their eyes across the Atlantic, or at least across the English Channel, when envisaging their future careers.

The dominance of the English language is a powerful factor, but it is not the sole reason for the net migration of scientific talent to North America and Britain. Also important are the highly individual academic career ladders that persist in different continental European countries. Employment rules and social security arrangements are also handled differently between countries. The fact that Euroland is such an academic patchwork tends to deter foreigners, and makes it difficult for natives who have gone abroad to return.

Some progress is being made. The European Commission has supported various programmes to promote the movement of young researchers between the European Union's member states. And in 1999, in Bologna, Italy, education ministers from 29 European countries agreed to harmonize aspects of their higher-education systems. These countries have pledged to introduce by 2010 comparable systems of Bachelors and Masters degrees, and a system of academic credits that will allow students to move between countries.

But despite these welcome initiatives, national isolation remains the rule, and no identifiably 'European' scientific career is in sight. It is no wonder that the continent's young scientists are crying out for more consideration to be paid to their plight (see Correspondence, page 259).

In countries such as Italy and Spain, for instance, the academic powers-that-be seem unwilling to reform recruitment systems that favour those who stay at home over — often stronger — candidates who have sought experience in foreign labs. Even programmes established with the goal of bringing back talented scientists working abroad have been blighted in this way (see *Nature* **413**, 556; 2001).

Elsewhere, attempts to address national idiosyncrasies have left some problems unsolved. Germany, for instance, the dominant producer of PhDs in Europe, has embarked on an academic reform aimed at creating better opportunities for young scientists (see News Feature, pages 257–258). But there are concerns that the reforms have not been adequately resourced, and also that they will leave stranded those researchers who have been employed on fixed-term contracts for more than 12 years. They must now find a permanent post, or leave academia.

The French government, meanwhile, recently opened up more tenured posts in response to fears about creating a 'lost generation' of researchers, who would not be available to replace the ageing baby-boomer generation. But it has not tackled the underlying problem of a system that gives postdocs no official social security status or contract rights, and so forces many French nationals wanting postdoctoral experience to go abroad (see *Nature* **414**, 145; 2001).

If Euroland's best young minds are to move between its nations with the same enthusiasm that currently lures them to North America and Britain, university administrators, immigration offices and society at large must learn that foreign researchers are not troublesome inconveniences, but highly motivated workers who offer their skills to the benefit of their hosts. ■

Welcome back to the fusion fold?

The United States should rejoin the ITER project, whose management must communicate better with politicians and public.

Governments and scientists worldwide will welcome indications that the United States is once again considering participation in ITER, the international project to build an experimental magnetic-confinement fusion reactor (see page 247).

If the United States does rejoin the project, it will send important political signals to the nation's allies, who are concerned about the unilateralist leanings of George W. Bush's administration. ITER is also the largest example of a research project pursued as a genuine collaboration between the world's major scientific powers. In most 'big science' initiatives, such as the US-led International Space Station, other partners have assumed a secondary role. But future projects in disciplines such as high-energy physics will require cooperation on equal terms, and ITER provides the only extant model.

ITER is also important because any real exploration of long-term, sustainable energy must include a thorough technical evaluation of nuclear fusion. Despite theoretical doubts about the ability of a doughnut-shaped magnetic chamber, or tokamak, to contain plasma, such a device remains the best prospect for tapping fusion power.

Before hopping back into bed together, however, the United States and its former partners need to take a careful look at what led to their 1999 estrangement.

Encouraged by the doubters in the US fusion-research community, Congress had lost patience with ITER. The project's management must share the blame, having failed to adapt its design proposal to fiscal realities — not just those in the United States, but those in Europe, Japan and Russia, too — until forced to do so. ITER also failed to sell itself to the public, as any big-ticket project must do in a democratic society.

Bush would do well to buy back into ITER, and to help to rebuild it as a model of scientific collaboration. The mercurial power of Congress to abruptly end funding will remain a threat — reforming the US budget process to provide some security of funding for the nation's international commitments may be just as challenging as containing plasma at a temperature of one million kelvin. In the meantime, ITER's management must do a better public-relations job, and ensure that it does not become distanced from its political paymasters. ■





End of the road
Loan trouble could
be the last straw for
chimp facility
p248



Power play
US health secretary
muscles in on
research agencies
p250



Sick leaves
Killer tree disease
sparks fresh concern
in California
p251



Brown dwarf
Image of star's
sidekick is an
astronomical first
p252

United States pledges to rethink rejection of fusion experiment

Geoff Brumfiel, Washington

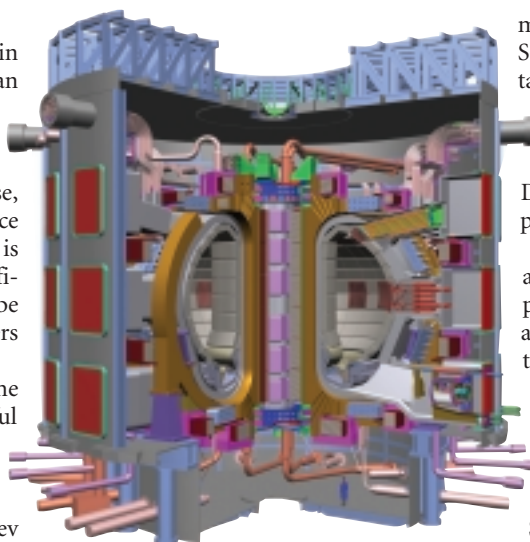
The United States may be poised to rejoin ITER, the international project to build an experimental magnetic fusion reactor, which it abandoned three years ago.

After several months of renewed interest in the project at the White House, John Marburger, President Bush's science adviser, confirmed that the United States is reviewing its involvement in ITER. "I definitely think that our participation should be reconsidered," he told a meeting of reporters on 8 January.

The United States, Japan, Russia and the European Union began designing a powerful new fusion experiment, then known as the International Thermonuclear Experimental Reactor, after US president Ronald Reagan and Soviet leader Mikhail Gorbachev endorsed the idea at a summit in 1988.

Fusion researchers want the energy emitted from ITER to surpass the energy put into it — a goal that has eluded them for over 50 years. After 10 years of planning, the team arrived at a design for a massive doughnut-shaped machine more than twice the size of any existing fusion device.

But ITER's projected \$10-billion price tag, together with technical concerns over the design, drained support in the United



Reactor revisited: the project's backers hope that the revised ITER will win back US support.

States, and in 1999, on instructions from Congress, the US Department of Energy withdrew from the project.

That hard-line stance now seems to be softening. In November, Sherwood Boehlert (Republican, New York) and Ralph Hall (Democrat, Texas), the chairman and senior

minority member, respectively, of the House Science Committee, wrote to energy secretary Spencer Abraham asking him to send an observer to this year's negotiations on ITER's construction. In his reply, Abraham promised to review the Department of Energy's stance on the project "in the next few months".

Project supporters hope that the Bush administration's general support for fusion power, its desire for visible political cooperation with its allies in the wake of 11 September, and revisions made to ITER's design since 1999, can combine to entice the United States back to the project.

After a reduction in the scale and scope of ITER, the project is now estimated to cost \$4.2 billion. The United States' reconsideration comes as ITER's participants begin negotiations on the facility's construction. Within the next year, the collaboration is due to select a site and decide how the costs should be divided up.

"The participants are working against a very tight schedule," says Jean-Pierre Rager, spokesman for the European negotiating team. If the United States were to rejoin, Rager encourages it to do so by June this year. "The more these negotiations advance, then the more difficult it is for

Protests mark threatened closure of medical school

Alison Abbott, Berlin

Clinical researchers have responded with alarm to a clause in the new Berlin state government's coalition agreement that pledges to close the medical faculty of the Free University and to remove university status from its Benjamin Franklin Clinic.

Over 4,500 researchers, clinicians, patients and members of the public demonstrated against the closure last Friday.

Since its reunification, Berlin has been home to two major research hospitals. The second is located at the Charité Hospital, and belongs to east Berlin's Humboldt University. Berlin's post-reunification

governments have already rationalized many of the duplicated departments (see *Nature* 380, 278; 1996). In its election campaign, the PDS — the reformed communist party, which has now formed a governing coalition with the Social Democrats — proposed a merger of the two faculties, but neither coalition party had previously suggested that either should be closed entirely.

Closure of the Free University's medical school would save Berlin 95 million euros (US\$85 million) per year, according to the new government. But Martin Paul, the faculty's dean, says that closure would

also mean the loss of 25 million euros that researchers bring in as external grants every year, as well as the jobs of 500 highly qualified staff.

The new science and culture minister, the PDS's Thomas Flierl, has publicly confirmed that he intends to proceed with the plan, although since the demonstration his party has indicated a willingness to consider arguments against closure.

Speculation that the decision to close the Free University's medical school is being driven by east-west rivalry has been fired by a lack of information about the coalition's true motive, according to Paul.



John Marburger thinks a US role in ITER "should be reconsidered".

Fusion researchers are enthusiastic about participation in ITER — provided that additional resources are made available to pay for it. "I think the community is very excited about the possibility of rejoining ITER," says Richard Hazeltine, director of the Institute for Fusion Studies at the University of Texas at Austin. Hazeltine criticized ITER before the US withdrawal, but thinks that the new design addresses his concerns.

The United States still has a long way to go before it could rejoin the project, cautions Anne Davies, director of the Office of Fusion Energy Sciences at the energy department. "We're just at the beginning stages of considering what our position should be," she says. Congress and the administration must pledge their full support before US fusion researchers could resume participation, she adds.

♦ www.iter.org

the United States to join," he says.

Japan and Canada have lobbied hard to get the Bush administration to give the project a second look. Canada, which joined the collaboration in 1997, is particularly keen on US participation, which it sees as lending credibility to its bid to host construction.

New-generation cars become old hat as US changes course

Mark Schrope

The Partnership for a New Generation of Vehicles (PNGV), a collaboration between the US government and industry to develop more efficient cars, is being replaced with a new programme called Freedom Cooperative Automotive Research, or Freedom CAR.

US energy secretary Spencer Abraham announced the change on 9 January at the Detroit Auto Show, saying that the aim is to "promote the development of hydrogen as a primary fuel for cars and trucks, as part of our effort to reduce American dependence on foreign oil". He added that the PNGV was not cost-effective and was not producing cars that were ready for the showroom.

Set up in 1993, the PNGV was strongly championed by then vice-president Al Gore. Its brief was to produce environmentally friendly prototype cars that were three times as fuel-efficient as those for sale at the time.

The programme included support for research into cars powered by hydrogen fuel cells. But the technology was not available to meet the 2004 deadline, and the PNGV turned to improving existing technologies.

Environmentalists are interpreting the announcement as a retreat from the Clinton administration's commitment to improving fuel economy in the short term. "This is a

decision to coddle the US automotive industry, rather than hold it accountable for making real improvements in vehicle efficiency over the coming years," says John DeCicco of Environmental Defense, a New York-based pressure group.

Vernon Roan, director of the fuel-cell lab at the University of Florida in Gainesville and vice-chairman of the National Research Council's peer-review panel for the PNGV, says that the programme had many successes. He notes that the PNGV helped to build bridges between the government and the car industry, produced useful prototypes, and expedited the plans of Ford and General Motors to build and sell fuel-efficient 'hybrid' vehicles that run on both petrol and battery power.

In its last review, the panel recommended substantial changes to the PNGV, and Roan says the new programme is consistent with these. "I think it is a good move," he says.

Although details of Freedom CAR have not been finalized, Mick Schwarz, who will lead Ford's participation in the programme, says that it may encourage the construction of infrastructure — including hydrogen filling stations. Schwarz says that he hopes the new programme's budget will match or exceed the \$127 million given to the PNGV this year.

♦ www.ta.doc.gov/pngv

Bank wants money back from troubled chimp facility

Erika Check, Washington

A controversial facility in New Mexico that houses chimpanzees for research is facing a serious threat to its survival.

The Coulston Foundation's facility at Alamogordo has weathered years of financial problems, regulatory troubles and animal-rights protests. But now a New Mexico bank has filed a foreclosure lawsuit against the foundation, which houses about 250 chimps. The bank alleges that Frederick Coulston and his foundation have defaulted on over \$1.16 million in loan repayments.

It is unclear how the foundation will deal with the lawsuit. A representative declined to comment, referring enquires to Coulston's lawyer, who did not return telephone calls.

But Joe Bielitzki, a member of the board of the Scientists Center for Animal Welfare (SCAW), a non-profit group that supports animal research and promotes the humane treatment of research animals, says that researchers are concerned about what will happen to the chimps if the foundation shuts down. "Everybody assumes that the National

Institutes of Health or somebody would pick up the ball and run with it, but the question is who has responsibility for these animals," he says. "It's a really sticky, ugly situation."

Over the past three years, the foundation has lost all of its funding from the National Institutes of Health (NIH). Last year, the NIH's National Center for Research Resources (NCRR) removed the chimps from Coulston's responsibility — although they are still held in the same compound at Alamogordo. In 1999, the NIH decided against renewing two other grants to Coulston.

In July, the US Department of Agriculture alleged that the foundation had violated the Animal Welfare Act before the deaths of two chimps. In October, the Food and Drug Administration refused to grant permits for new experiments by the foundation.

The foundation has also long been a target of animal-rights protesters. Last September, the Animal Liberation Front claimed responsibility for a fire at Coulston's facility.

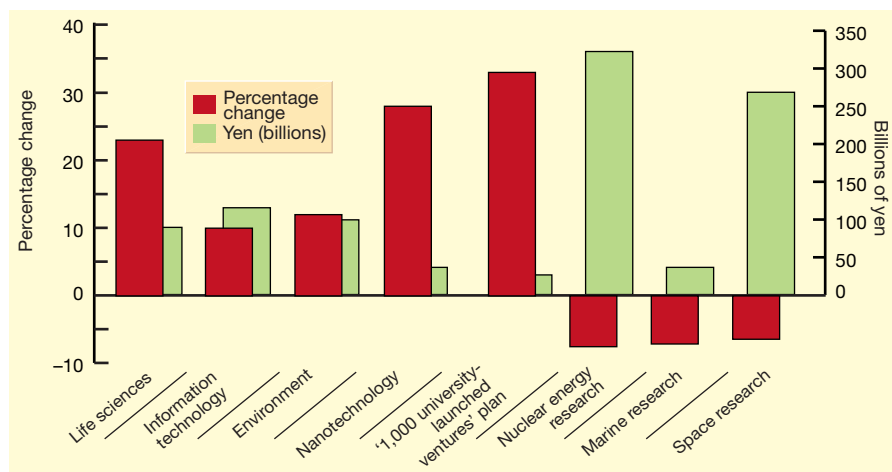
Space for chimps in the United States is limited, and the issue of their long-term care



Down and out: researchers are worried that the Coulston chimps could end up homeless.

is problematic. Chimps are now rarely used in research, and this year the NCRR will award a \$5-million contract for the construction of a 'sanctuary' to house about 200 retired chimps. However, this will not be ready for several years, and the NIH has already allocated animals to it.

Japan sets focus on critical areas and economic pay-off



A question of priorities: proposed spending for the next year will bring winners and losers.

David Cyranoski, Tokyo

Japan may be mired in a recession, but the slump won't stop its efforts to compete internationally in fields of research that it deems to be strategically important.

That is the central message of the budget for the fiscal year starting in April, which was released last month. The budget boosts research in nanotechnology, environmental sciences, information technology and areas of the life sciences that have been designated as national priorities (see graph).

But with the overall research budget growing by only 2%, other fields, among them marine science, nuclear energy and space research, will suffer cutbacks.

And the budget comes with strings attached, as the government tries to make sure that its spending on research yields economic benefits. The Ministry of Education, Culture, Sports, Science and Technology (MEXT), for example, is to set up a team of patent lawyers and researchers whose remit is to ensure that the protein research supported by MEXT yields new patents.

"It's not something we're happy about," says a researcher at the nuclear magnetic resonance facility in Yokohama, which is run by the Institute of Physical and Chemical Research (RIKEN). "But that's how the money comes." Another researcher there adds: "It's a lot of pressure."

Researchers are increasingly aware of the expectations for them to produce applicable research. MEXT and the Ministry of Economy, Trade and Industry plan to invest ¥248 billion (US\$2 billion) — an increase of 28% on this year — in technology transfer. Their plan includes a target of creating 1,000 start-up companies over the next three years, based on university research.

Among projects that are affected by cuts in

the budget is the Institute of Space and Astronautical Science's Venus mission, which aims to pierce the planet's thick clouds with infrared lasers to investigate its atmospheric conditions and volcanic activity. It will get only 10% of the ¥1 billion requested by the institute, and its planned 2007 launch is set to be deferred. "Existing projects and those with international commitments are protected, but it will be very difficult to start new ones," says the institute's deputy director, Toshio Matsumoto.

Several research institutes will also pay the penalty for their status as 'special public corporations' (see *Nature* 414, 833; 2001). The government intends to restructure these, and is imposing budget cuts of around 10%. "The government is boasting of a budget increase for science, but that excludes us just because we are a 'special public corporation'. It's incomprehensible to me," says a senior researcher at the Japan Marine Science and Technology Center in Yokosuka. Operation of a new ocean-drilling ship, due to begin in 2006, will probably be delayed as a result.

RIKEN, another 'special public corporation', will have its budget cut by 9%. Major projects, including a proposed radioisotope-beam facility at Wako, northwest of Tokyo, will be the main casualties. RIKEN vice-president Tomoya Ogawa says that the facility will probably be delayed by "a few years".

Some research administrators are accepting such reductions more stoically than others. Japan's High Energy Accelerator Research Organization (KEK) will get extra money to build a 50-GeV particle accelerator, whereas the rest of its budget will fall by 5% — after a 10% drop this year. But KEK director Hirotaka Sugawara claims that the cut will not damage research, and that he plans to hire staff. "We expect a wonderful year," he says. ■

Space agency pulls the plug on astrometry mission

Tony Reichhardt, Washington

A NASA mission that was to have measured the positions of stars with unprecedented accuracy has been cancelled because of cost overruns and problems with its detectors.

The Full-sky Astrometric Mapping Explorer (FAME), which was selected in 1999 as one of the agency's medium-sized Explorer missions, was due to launch in 2004. It would have mapped the positions of 40 million stars with 20 times the accuracy of Europe's Hipparcos mission, the best existing astrometric survey undertaken from space.

NASA wrote to FAME's principal investigator Kenneth Johnston, of the US Naval Observatory in Washington DC, on 4 January, saying that it was withdrawing its support because of growth in the project's estimated total budget from \$180 million to \$220 million. The agency also cited difficulties in obtaining charge-coupled device (CCD) detectors for the mission. The manufacturer, Scientific Imaging Technologies of Tigard, Oregon, has had trouble producing CCDs to the required specifications, Johnston says.

Project managers had already scaled back the mission, and Johnston offered to try to obtain further funds from the Navy to ease the cost overrun, but NASA saw too much risk in continuing with the mission. "There's a real problem at NASA headquarters with budgets," Johnston says. "They have to do something. We just happened to be in the wrong place at the wrong time."

FAME's cancellation leaves a gap in astrometry missions. NASA's more sensitive Space Interferometry Mission will not be launched until 2009, and Europe's proposed GAIA mission is itself in jeopardy and in any case is not scheduled to begin until 2011 at the earliest (see *Nature* 414, 383; 2001). Robert Reasenberg of the Harvard-Smithsonian Center for Astrophysics, a former project scientist for FAME, laments the loss of the science, but agrees that NASA has to take action to rein in projects that are running over budget. "If I were starting a programme now, this decision would loom pretty large," he says.

Johnston confirms that the mission had problems procuring the CCDs. "CCDs sound like they're off the shelf," he says, "but it's not a guaranteed process."

Johnston still hopes eventually to secure funding from NASA or the Navy, although he admits that it's a "big mountain to climb". ■

Health secretary under attack over centralization plan

Erika Check, Washington

Tommy Thompson, the US health secretary, is under heavy fire this week for his far-reaching plan to streamline public and Congressional relations at the National Institutes of Health (NIH) and other parts of his sprawling department.

Thompson wants enquiries to all parts of the Department of Health and Human Services (HHS) — including the Food and Drug Administration, the Centers for Disease Control and Prevention (CDC) and all the institutes of the NIH — to pass through new, centralized departments under his control.

Health-department officials say that the reorganization will help them to present clearer, more coherent messages on crucial science and health matters. "We've been working since we got here to try to create a more unified department and a more responsive structure," says Kevin Keane, assistant secretary for public affairs at the HHS.

But supporters of the various agencies are deeply concerned that the change would damage the institutes' credibility with the public and Congress. "Information from the NIH has been highly valued for many years because it is viewed as a bipartisan institution with no political agenda," says Harold



Tommy Thompson wants to see a more unified structure.

Varmus, former director of the NIH.

The restructuring is expected to form part of the budget proposal for the 2003 fiscal year, which will be sent to Congress by President Bush in early February. Strong objections are expected, and Congressional staff are already on the attack. "We need free and unfettered access to the various agencies," says one Democrat aide in the Senate.

Earlier efforts by Thompson to direct press and public enquiries through his office caused problems last autumn, when CDC and NIH officials were told to steer questions on anthrax through it. As a result, critics claim, the public did not receive accurate and timely information during a series of anthrax attacks.

But Keane claims that fears over restricted access to scientific information are unfounded. "We're just trying to handle communications efficiently so there will be a clear place to turn to for what people need, and a clearer answer coming back," he says. ■

Allegations of police surveillance prompt calls for inquiry

Sally Goodman, Paris

INSERM, France's main biomedical research agency, has been asked to investigate allegations that one of its laboratories came under police surveillance because of its work on the role of dietary salt in heart disease.

The allegations, which have been emphatically denied by the French authorities, appeared in the 11 January edition of *Le Point*, one of France's leading weekly news magazines.

The magazine says that in May 2000, the police received an order from an unknown source to tap the telephone of Pierre Meneton, a researcher at INSERM's laboratory of physiology and experimental vascular pathology in Paris. That was shortly after Meneton handed a report to the French food-safety agency, AFSSA, in which he accused the food industry of "poisoning" consumers because of excess salt in its products.

Meneton was surprised by the allegations in *Le Point*. "I will be asking for a full investigation and if there is any truth in the matter I will be lodging an official complaint," he says.

The relationship between salt intake and high blood pressure has long been the subject of fierce scientific argument. Publication of the allegations coincided with an international conference in Paris on salt and health, organized by the AFSSA in response to Meneton's report. An advisory panel set up by the agency told the meeting that the salt content in processed food should be cut by 20% over five years.

Le Point claimed that a member of the panel, Joël Ménard, a former chief health officer in the French government, has also been under police surveillance. But Ménard is sceptical. "I can't believe it's true," he says. "If there is any truth in it, then French democracy is dead." ■

Astronomer set to star in Taiwan

David Cyranoski, Tokyo

Taiwan has enticed Frank Shu, a leading astronomer in the United States, to take up the post of president at the National Tsinghua University in Hsinchu.

Shu, who is based at the University of California, Berkeley, is a prominent theorist and former president of the American Astronomical Society.

Taiwanese government officials see Shu's



Frank Shu sees scope for growth at Tsinghua in Taiwan.

appointment as a significant coup for its university system. Shu says that his move is motivated by one basic assessment: "I thought I could make a bigger difference in Taiwan."

Working with C. C. Lin at the Massachusetts Institute of Technology, Shu in 1964 proposed the spiral density wave

theory, which provided the first quantitative model for understanding how spiral galaxies form. He went on to examine the interstellar medium, planetary rings and, more recently, star formation and meteorites.

"The proper role of the university president is to serve faculty and students," Shu says, pledging to continue his research at Tsinghua, "and you can only have the necessary understanding if you are down in the trenches yourself." He aims to make the university bilingual, with courses offered in Chinese and English.

Shu's move "will help to recruit better talent to Taiwan and Asia," says Fred Lo, director of the Institute of Astronomy and Astrophysics in Taipei. The appointment "will also have influence beyond the university campus and beyond astronomy," he adds.

Born in China, but raised in the United States, Shu is following a precedent. At 58, Shu's current age, his father took over the presidency of Tsinghua university and helped to develop the neighbouring Hsinchu Science-based Industrial Park that gave rise to Taiwan's electronics industry. ■



Fears mount as oak blight infects redwoods

Jonathan Knight, San Francisco

A disease that is killing oak trees by the thousand in California has been found in the dying needles of the coast redwood (*Sequoia sempervirens*). The redwood is the world's tallest tree and has great symbolic and economic importance for the state.

Tests to determine whether the pathogen infects healthy redwoods, or is merely an opportunist that attacks already ailing trees, are now under way at a University of California nursery near Monterey, and may be completed within weeks.

But even if the fungus-like organism that causes the disease proves to be no great threat to redwoods, it is rapidly becoming one of the most damaging forest pathogens to have emerged in the United States in recent years.

In 1995, large numbers of tanoak (*Lithocarpus densiflorus*) and coast live oak (*Quercus agrifolia*) began dying in counties surrounding San Francisco Bay. The phenomenon was called sudden oak death (SOD) because once the first symptoms had appeared — wilted shoots and a deep red sap oozing from the bark — the entire canopy would turn brown within two or three weeks.

SOD has now spread to many areas within a 300-kilometre swath of coastal forest. In June 2000, David Rizzo, a plant pathologist at the University of California, Davis, found that the likely culprit was a member of the genus *Phytophthora*, to which potato blight also belongs.

Ridge-top infection

Although similar symptoms appear in trees infected with known *Phytophthora* species, the epidemiology was unusual, Rizzo says: "The infected trees were in a weird place for a *Phytophthora*. They were on ridge tops, not near streams."

DNA sequencing in the laboratory of Matteo Garbelotto, a plant pathologist at the University of California, Berkeley, revealed that Rizzo's *Phytophthora* was new to California but had previously been identified in rhododendrons in Germany and the Netherlands. Researchers at the Federal Biological Research Centre for Agriculture and Forestry in Braunschweig, Germany, first described the species, and named it *P. ramorum*.

To show that *P. ramorum* was the killer, Rizzo and Garbelotto experimentally infected oak seedlings at the university's nursery near Monterey, and found the same symptoms. They also inoculated the bark of healthy oak trees in Marin county, north of San Francisco, in an area where SOD was already prevalent. Within weeks, massive 'bleeding' lesions had grown that ultimately girdled the trees.

Unlike Dutch elm disease, SOD does not restrict itself to a single type of tree.



Portrait of a disease: Californian redwoods (top) may be falling victim to the fungus-like pathogen *Phytophthora* (left). Symptoms include brown leaves and oozing red sap.

The California Department of Forestry, researchers and even members of the public have reported that trees other than oaks have the same symptoms as SOD. So far, Rizzo and Garbelotto have confirmed that *P. ramorum* causes disease in 15 different species, including huckleberry, buckeye, madrone and bigleaf maple.

Last autumn, researchers at the University of California, Berkeley, noticed that the stems and needles on several of the redwoods on campus suddenly turned brown. Garbelotto's laboratory isolated *P. ramorum* DNA from the needles using the polymerase chain reaction.

"The shoots are symptomatic," says Garbelotto, "but it's a big jump to say that the tree has the disease." He is now testing *P. ramorum* infectivity in redwood seedlings. So far, he says, every species inoculated has been susceptible.

Damage limitation

Even if redwoods are a host for *P. ramorum*, the infection may not be lethal. SOD usually kills oaks and sometimes rhododendrons, but in other species the infection just causes leaf spots or the death of some shoots without killing the tree.

So far, *P. ramorum* has not done as much damage as some of its close relatives. *P. cinnamomi*, for example, has reduced vast tracts of the Jarrah forest in Western Australia to grassland since it was accidentally introduced in the 1920s. It infects the roots of nearly a thousand different species, and can be lethal in many of these if the soil is soggy,

says Everett Hansen, a forest pathologist at Oregon State University in Corvallis.

But unlike most other forest *Phytophthora* species, *P. ramorum* spreads through the air, rather than through the soil. "It is something we have not had to contend with before in forest pathology," Hansen says.

Garbelotto has found very few genetic differences between different isolates of *P. ramorum* from California, suggesting that it is a recent introduction. So far the only way of stopping the disease is to cut and burn infected trees, but Garbelotto's group is testing a variety of chemicals on laboratory cultures to see if any are effective against the pathogen. ■

♦ www.suddenoakdeath.org



Off colour: the diseased redwoods' condition resembles sudden oak death, in which the entire canopy can turn brown in a matter of weeks.

J. KNIGHT

FAR LEFT: GARBOLETTO LAB, UC BERKELEY; LEFT: D. RIZZO

J. KNIGHT

Smithsonian lobby forces White House to abandon reform plans

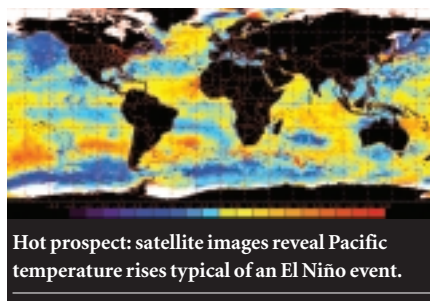
Washington Scientists at the Smithsonian Institution are quietly celebrating the defeat of a proposal to transfer the management of its best-known research facilities to the National Science Foundation (NSF).

Under a plan hatched by the White House's Office of Management and Budget, the presidential budget request for 2003 was expected to shift \$35 million to the NSF, making the agency responsible for the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, the Smithsonian Tropical Research Institute in Panama, and the Smithsonian Environmental Research Center in Edgewater, Maryland (see *Nature* 414, 680; 2001).

Sources at the Smithsonian say that six members of its board of regents went to the White House to argue against the decision, saying that the transfer would cost more money and that research funding is already very competitive. The White House is now expected to abandon the proposal, requesting instead a study of how to support research in the Smithsonian.

Pacific warming raises fears of another El Niño

Washington Nations that fringe the Pacific Ocean could be struck by another El Niño event within months, according to the US National Oceanic and Atmospheric Administration (NOAA). The agency's satellites have detected a warming of the



tropical Pacific that is characteristic of the phenomenon, which typically disrupts the climate on both sides of the ocean.

The last El Niño event, which was particularly powerful, subsided in 1998. "At this point it is too early to predict if this El Niño might develop along the same lines as the 1997–98 episode or be weaker," says Vernon Kousky of the NOAA Climate Prediction Center in Washington.

French researchers hit by agency age change

Paris France's main research agency, the CNRS, has upset the career plans of hundreds of young researchers by reducing the age limit for those applying for a permanent position. The move comes only a month before the deadline for applications.

The change reduces the age limit for the lowest level of entry on the CNRS research career ladder from "under 32" to "under 31", and brings it in line with other public research agencies. Last year, over 400 candidates between these ages applied for positions at this grade. Around 80 were successful.

CNRS director-general Geneviève Berger admits that the situation is "painful" and is

creating up to 80 two-year postdoctoral positions in 2002 for the affected candidates. She is also encouraging them to apply for higher-grade posts, which have no age limit.

Dispute over stem-cell rights is resolved

Washington A squabble over who can work on the world's first human embryonic stem (ES) cell lines has been resolved. James Thomson of the University of Wisconsin at Madison unveiled the first human ES cells in 1998. But since last August, the company that backed his work, Geron of Menlo Park in California, has been in dispute with the Wisconsin Alumni Research Foundation (WARF). The foundation challenged Geron's claim to exclusive rights to develop therapies based on the differentiation of the cells into a variety of cell types (see *Nature* 412, 753; 2001).

The new agreement gives Geron exclusive rights to develop therapies and diagnostics based on nerve cells, cardiac muscle cells and pancreatic islet cells, which secrete insulin, and non-exclusive rights to some other cell types. WARF says it will allow academic researchers and government scientists to work on the cells "without royalties or fees".

US shifts its stance on nuclear weapons

Washington The US administration has unveiled a strategic plan to maintain its stockpile of nuclear weapons while reducing the number of warheads deployed.

The number of US warheads in commission will be cut from 6,000 to roughly 2,000 over the next decade. But none of the decommissioned warheads will be destroyed, and officials say that an unspecified number will be put into storage. The plan also calls on the US Department of Energy to reduce the time needed to prepare for the resumption of underground bomb testing — although a 1992 US moratorium on testing will remain in place for now.

The plan has been attacked by arms-control experts and the Russian government.

Yucca Mountain chosen as radioactive waste dump

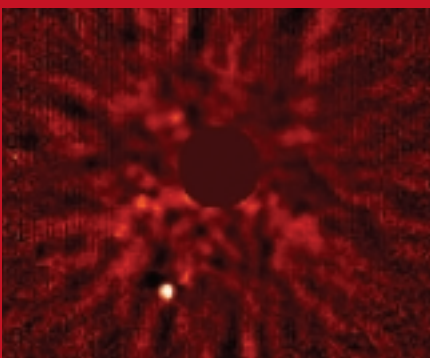
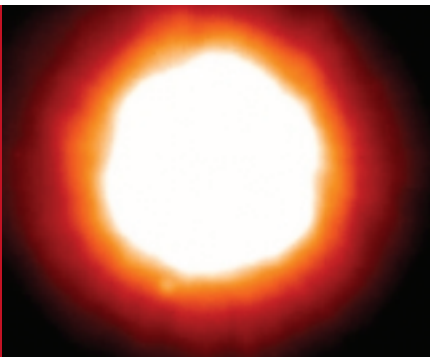
Washington Nevada's Yucca Mountain should be the site of a long-term repository for nuclear waste, US energy secretary Spencer Abraham announced last week.

The proposed underground repository will house up to 77,000 tonnes of nuclear waste. If, as expected, President George W. Bush gives his approval, the state of Nevada will have 60 days to notify Congress of its disapproval. Congress can overturn Nevada's objection by a simple majority vote, but any subsequent decision to go ahead is likely to be challenged in court (see *Nature* 415, 6; 2002).

Brown dwarf shines to put astronomers in orbit

Washington The brown dwarf shown here orbiting the star 15 Sge is the closest object yet photographed orbiting a Sun-like star outside the Solar System. Michael Liu of the University of Hawaii and his colleagues took the image using adaptive optics, which correct for atmospheric distortion, on the 8.1-metre Gemini North telescope in Hawaii. They unveiled it last week at the American Astronomical Society's meeting in Washington.

The lower image has been processed to suppress the light from the star, so the brown dwarf can be clearly seen near the bottom. The object orbits 15 Sge at 14 times the distance of the Earth from the Sun — which in our Solar System would be between Saturn and Uranus. Brown dwarfs are intermediate between planets and stars; they are not massive enough for nuclear fusion to occur in their cores.



Sexual stereotypes

Males are promiscuous and females are choosy, according to evolutionary dogma embodied in a theory called Bateman's principle. Only recently have researchers begun to test the theory's limits, says Jonathan Knight.

*"The female, with the rarest exceptions, is less eager than the male ... she is coy, and may often be seen endeavouring for a long time to escape."*¹

With these words, Charles Darwin cast the die for evolutionary studies of male and female sexual roles. Darwin realized that the peacock's tail and the lion's mane owe their existence to aeons of fierce competition between their male ancestors for mates. Competitive males, Darwin assumed, will attempt to mate at every opportunity; females, he reasoned, are inherently choosy, reserving their favours for the strongest or gaudiest suitor.



The mating game: the image of coy females and ardent males is an old one. Bateman's principle argues that these roles stem from the economics of sperm and egg production (main picture).

These ideas were later crystallized in a theory known as Bateman's principle. In 1948, the British biologist Angus John Bateman concluded from experiments with fruitflies that promiscuity is more valuable to the reproductive success of males than to that of females. Males, he concluded, have therefore evolved an "undiscriminating eagerness" to mate, whereas females display "discriminating passivity"² — a fundamental dichotomy that Bateman suggested even applies, to an extent, to our own species.

The explanation, Bateman argued, is that sperm are small and cost next to nothing to produce — so the wider a male can spread them, the better off he will be. A female, on the other hand, produces many fewer eggs, and invests a relatively large amount of energy in each one. All she really needs is one good male to fertilize them to reach her maximum reproductive output. It all seemed to make sense, and Bateman's principle soon became one of the grounding truisms of behavioural biology.

But several researchers have since taken exception to this characterization of male and female sexual roles. Objections were first heard, but largely ignored, in the 1970s. Today, behavioural biologists are finding evidence that the world of sex is more complicated than Bateman thought. It's not that his principle is invalid, they say, but rather that it has been used to extend dated preconceptions about human sexual behaviour to

the entire animal kingdom, sometimes to the detriment of scientific knowledge.

Despite the Victorians' reputation for prudishness, nineteenth-century natural philosophers spent lots of time watching animals mate. At the height of the British Empire, there were few places in the world where courting animals could escape the note-taking naturalist, and Darwin relied heavily on these descriptions in developing his theory of sexual selection¹, with its underlying assumptions of eager males and reluctant females.

Counting on success

Bateman, who worked at the John Innes Horticultural Institution near London, was one of the first biologists to furnish the theory with experimental evidence. In a series of genetic experiments², Bateman combined groups of male and female fruitflies (*Drosophila melanogaster*) in vials. Each fruitfly carried a different dominant genetic marker, so Bateman could score the breeding success of each individual by counting the number of times the marker showed up in the next generation.

His conclusion was that, for males, promiscuity pays off. Males that mated with several females produced three to four times as many offspring as their monogamous peers. But females gained little by playing the field — on average, their reproductive output less than doubled when their progeny were sired by more than one male.

Bateman couched his explanation in terms of the relative energetic costs of produc-



Fly guy: Angus John Bateman (inset) used *Drosophila* to study the benefits of promiscuity.

MEHAU KUIYK/SPL; INSET: CORBIS

CLAUDE NURDSANY & MARIE PERENNOU/SPL; INSET COURTESY OF THE JOHN INNES FOUNDATION TRUSTEES

ing sperm and eggs, but today the principle that bears his name includes an important modification made in 1972 by Robert Trivers of Harvard University, one of the founders of the discipline of sociobiology³. He expanded Bateman's concept beyond gametes to include a parent's entire investment in its offspring — including gestation, feeding and protection. The sex that invests more, Trivers argued, should be more passive and discriminating, whereas that with the smaller investment should court more mates and be ready to fight over them. This competition in turn drives Darwin's sexual selection by favouring traits such as showy colours and aggressiveness.

As evidence, Trivers cited well-known examples of sexual role reversal, such as seahorses and pipefish, and birds such as the greater painted snipe (*Rostratula benghalensis*). In this species, males incubate the eggs and rear the young. As the Bateman principle predicts, female painted snipes have brighter coloration and aggressively court males.

But over the next few years, a handful of primatologists began to notice behaviour in apes and monkeys that did not fit the Bateman mould. When doing field research in India on Hanuman langur monkeys (*Presbytis entellus*), Sarah Hrdy, then a graduate student at Harvard University, observed that, around the time of ovulation, females would aggressively seek copulations with multiple males. And when a new male rose to power in a group of langurs, even pregnant females would copulate with him.

Although female langurs are promiscu-

ous, this is not a classic case of sexual role reversal — the females still take on most of the burden of parental care. Hrdy, who is now an emeritus professor at the University of California, Davis, proposed an alternative explanation⁴. In langurs, as in several other primate species, males sometimes kill the unweaned young of females with which they mate, presumably so that more attention will be lavished on their own offspring when they are born. As a defence against infanticide, Hrdy argued, females may confuse the issue of paternity through flagrant promiscuity.

Hrdy's idea that promiscuity might have an adaptive value for females met with a great deal of criticism at the time⁵. Hrdy and her supporters became characterized as 'feminist' behavioural ecologists, and were seen by many researchers as a marginal group. But over the years, the number of cases of female promiscuity described in the animal-behaviour literature has mushroomed, particularly since the advent a decade ago of techniques for determining paternity by DNA analysis.

A degree of female promiscuity now seems to be the rule rather than the exception⁶. It has been documented in animals as diverse as whales⁷, rodents⁸ and bees⁹. In some animals, such as pseudoscorpions — arthropods that are related to scorpions and spiders — females prefer to switch partners much more often than males¹⁰.

Bateman would not have predicted that what is good for the gander is good for the goose. And behavioural biologists are now striving to understand what females gain from promiscuity. According to Hrdy's 'confused paternity' argument, a promiscuous female stands to get more help, or at least less interference, with raising her young. In certain insect species, where males often make nuptial gifts of food items, the benefits

Role reversal: the male frigatebird (left) shows off to impress the ladies, but in the painted snipe it is a different story — males (below left) rear the young and are outshone by females (below).



Sarah Hrdy (inset) argued that female langurs' promiscuity helps to safeguard their offspring.

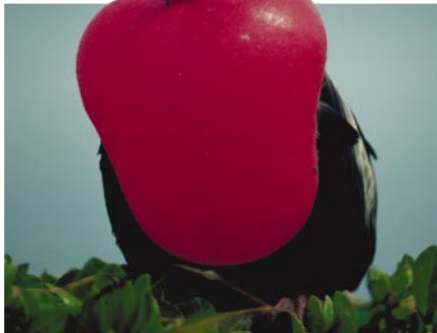
may be primarily nutritional¹¹.

But several recent studies have suggested that there may also be a benefit to the health of the offspring. John Hoogland of the University of Maryland's Center for Environmental Science in Frostburg has found that female prairie dogs (*Cynomys gunnisoni*) that mate with multiple males end up with larger litters, and that their pups are more healthy⁸. Female sand lizards (*Lacerta agilis*) and adders (*Vipera berus*) also do better with multiple mates¹².

Fishing for genes

Why should this be? Evolutionary biologists Jeanne and David Zeh of the University of Nevada in Reno argue that promiscuous females gain a genetic benefit: they improve their chance of finding a genetically compatible male. In 1999, the Zehs published an experiment¹³ that confirmed this principle in the pseudoscorpion *Cordylochernes scorpioides*. They allowed females to mate either twice with one male or once each with two males. Even though all females received the same amount of sperm, females that mated with multiple males had a higher percentage of offspring mature to adulthood. The Zehs argued that promiscuity can help a female by reducing problems with low offspring viability that can result from mating with a partner whose genes don't happen to complement her own. In essence, promiscuity means that a female does not have to put all her eggs in one basket.

Earlier this month, Tom Tregenza and Nina Wedell of the University of Leeds, UK, reported that female field crickets (*Gryllus*



► *bimaculatus*) may use promiscuity to avoid the cost of inbreeding¹⁴. In their experiment, females were allowed to mate with two brothers, with a brother and an unrelated male, or with two unrelated males. The second and third groups had similar reproductive success, but eggs laid by females in the first group were less likely to hatch. By some unknown mechanism, females that mate with multiple males seem able to fertilize their eggs selectively with the sperm of unrelated males, and so avoid the problems of poor embryo viability that are a consequence of inbreeding.

Multiple orgasms

The flip side of the Bateman equation — that males are boundlessly eager to mate because of their almost inexhaustible sperm supply — has also come under scrutiny. Although it may take very little energy to make a single sperm, males never deliver just one at a time. And in certain circumstances, the energetic costs of fertilizing a single female's eggs can be high.

Again, dissenting voices were first raised against this facet of the Bateman principle in the 1970s. Donald Dewsbury, a comparative psychologist at the University of Florida in Gainesville, noticed that male deer mice, hamsters and other rodents tended to copulate several times with the same female. It turned out that females are more likely to release the hormones needed for pregnancy if they are stimulated by multiple ejaculations, regardless of the amount of sperm they receive¹⁵.

As a result, Dewsbury pointed out in a 1982 review¹⁶ that a hit-and-run mating strategy might not always be the best option for males. If a male mates once and takes off, leaving the female in question open to the attentions of rival males, his sperm may not fertilize her eggs, and he also has to spend time and energy searching for another mate. Particularly when the number of receptive females is relatively low, a male may be better off sticking to one mate, and guarding her against rival suitors. This is a complexity that Bateman failed to recognize, Dewsbury argued. Under such conditions, he concluded, males might well be the choosy sex.

Bateman didn't actually report on the choosiness of his fruitflies, nor on any other aspect of their behaviour, notes Patricia Gowaty, an evolutionary biologist at the University of Georgia in Athens. "I think the reason Bateman's observation became 'Bateman's principle' is that it appealed to people's intuition about the behaviour of individuals," she says. "Very few people actually questioned the basic statement about the ubiquity of coy females and competitive males."

But if the layers of behavioural complexity are removed, how well does Bateman's basic theory about the relative benefits of promiscuity for the two sexes hold up? To



The Oregon newt is at the centre of a current study to evaluate Bateman's principle in the wild.

find out, several groups are now putting Bateman's principle to more rigorous tests.

Species that exhibit sexual role reversal are particularly useful for this purpose. In the pipefish *Syngnathus typhle*, for example, the males carry the young, whereas the females compete with one another for mates. In 2000, researchers at the University of Georgia and at Uppsala University in Sweden showed that the relationship between the number of mates taken and the number of offspring produced, which they termed the Bateman gradient, was also reversed. Taking multiple mates was of greater value to the fertility of females than to that of males¹⁷.

These experiments were done with captive fish. But the study's lead author, Adam Jones, who is now at Oregon State University in Corvallis, is attempting for the first time to measure Bateman's gradient in a natural setting — this time in a species with conventional male and female roles. He is gathering data from a local population of Oregon newts (*Taricha granulosa*), which conveniently breed in large aggregations so that every parent and offspring can be counted. The results are preliminary, but Jones says that so far they fit the Bateman model.

Cricket supporter

Leigh Simmons, of the University of Western Australia near Perth, meanwhile, has found empirical support for Bateman's principle in the bushcricket *Requena verticalis*. In this species, the male donates a droplet of nutrients to his mate. This is no problem when there is lots of food around. But under starvation conditions, males become reluctant to mate and females must compete for them. Simmons has shown in the laboratory that, when sexual roles are reversed in this way, the Bateman gradient is also reversed¹⁸. This, he argues, indicates that the basic Bateman principle is sound, despite the overlying layers of complexity. "Bateman was naively simplistic, but then it was 1948," Simmons says.

"The basic principle is right," agrees Tim Clutton-Brock, a behavioural ecologist at the University of Cambridge, UK. Females can gain from promiscuity, but the fact that they

more often get left holding the baby means that males can usually take more mates in a given time period¹⁹.

Although the principle remains valid, recent experiments have revealed intriguing complexities — even among Bateman's own experimental subjects. In 1999, researchers at University College London and Cornell University in Ithaca, New York, described proteins in fruitfly seminal fluid that increase the time that a female waits before allowing another male to copulate with her²⁰.

Gowaty likens the proteins to a chemical chastity belt. "Everybody assumes Bateman was about coy females and ardent males," she says. "Now here's a modern discovery that suggests maybe the reason females were holding back from mating is that they were being manipulated by a male protein."

This theory has yet to be tested. But if Gowaty is correct, Bateman's experimental observations may have had much less to do with inherent female coyness than he assumed — another example, perhaps, of the truth being obscured by nineteenth-century sexual stereotypes. ■

Jonathan Knight writes for *Nature* from San Francisco.

1. Darwin, C. *The Descent of Man, and Selection in Relation to Sex* (John Murray, London, 1871).
2. Bateman, A. J. *Heredity* **2**, 349–368 (1948).
3. Trivers, R. L. in *Sexual Selection and the Descent of Man* (ed. Campbell, B. J.) 136–177 (Aldine, Chicago, 1972).
4. Hrdy, S. B. *The Woman that Never Evolved* (Harvard Univ. Press, Cambridge, Massachusetts, 1981).
5. Symons, D. Q. *Rev. Biol.* **57**, 297–298 (1982).
6. Birkhead, T. R. & Møller, A. P. (eds) *Sperm Competition and Sexual Selection* (Academic, New York, 1998).
7. Clapham, P. J. & Palsboll, P. J. *Proc. R. Soc. Lond. B* **264**, 95–98 (1997).
8. Hoogland, J. *Anim. Behav.* **55**, 351–359 (1998).
9. Moritz, R. F. A. et al. *Behav. Ecol. Sociobiol.* **37**, 357–363 (1995).
10. Zeh, J. A., Newcomer, S. D. & Zeh, D. W. *Proc. Natl Acad. Sci. USA* **95**, 13732–13736 (1998).
11. Simmons, L. W. & Parker, G. A. *Ethology* **81**, 332–343 (1989).
12. Olsson, M. & Madsen, T. J. *Hered.* **92**, 190–197 (2001).
13. Newcomer, S. D., Zeh, J. A. & Zeh, D. W. *Proc. Natl Acad. Sci. USA* **96**, 10236–10241 (1999).
14. Tregenza, T. & Wedell, N. *Nature* **415**, 71–73 (2002).
15. Dewsbury, D. A. J. *Comp. Physiol. Psychol.* **93**, 178–188 (1979).
16. Dewsbury, D. A. *Am. Nat.* **119**, 601–610 (1982).
17. Jones, A. G., Rosenqvist, G., Berglund, A., Arnold, S. J. & Avise, J. C. *Proc. R. Soc. Lond. B* **267**, 677–680 (2000).
18. Kvarnemo, C. & Simmons, L. W. *Anim. Behav.* **55**, 1499–1506 (1998).
19. Clutton-Brock, T. H. & Vincent, A. C. J. *Nature* **351**, 58–60 (1991).
20. Neubaum, D. M. & Wolfner, M. F. *Genetics* **153**, 845–857 (1999).

Breaking the *Habilitation* habit

Germany plans to reform its antiquated academic career structure. But has the new model been fully thought through and is it adequately resourced? Quirin Schiermeier considers the evidence.

Madelaine Böhme throws heart and soul into her work as a palaeontologist, but she has no illusions about the career opportunities in her field. At 34, she wants to secure a tenured position. But in all probability, no suitable professorship will become vacant throughout her native Germany until 2007.

Böhme decided to stay at her bench at the Ludwig-Maximilian University of Munich because she loves her work. But she condemns the peculiarities of an academic system that prevent her from applying for a post outside the narrow confines of a traditional geology department. "Our institutional structures are a nineteenth-century fossil," Böhme complains. "Take evolution: there is independent



University challenge: can Edelgard Bulmahn (above, left) inspire Germany's young minds?

work going on in the biology and geology departments, but there is no interdisciplinary unit for evolutionary research."

In November, Böhme and a round table of 11 other young scientists, plus a handful of senior researchers, funding-agency officials, and editorial staff from *Nature*, gathered in Munich to discuss the strengths and weaknesses of the German academic system, and the prospects for reform. Most agreed that changes are needed. But doubts emerged about the specific plans being put in place by science minister Edelgard Bulmahn. In particular, many attendees were worried that the envisaged reforms — although an improvement on the status quo — impose a rigid timetable on academic careers and fail to match the support offered under the US tenure-track system.

Qualified criticism

Complaints about Germany's current university career structure have been piling up on Bulmahn's desk for some time. Last year, she announced a reform package designed to overhaul the system and make it more accessible and attractive to the best young minds. At the heart of these changes is a plan to replace the *Habilitation* qualification. Unique to German-speaking countries, the *Habilitation* is based on a post-PhD second research thesis, and is usually required for a professorship in a given academic field.

Most young scientists see the *Habilitation* as an anachronism that can keep researchers dependent on their professors into their forties. In many cases, the restrictions it imposes last even longer. A *Habilitation* awarded in one department is usually only recognized by oth-



ers in the same discipline. Because Böhme will have the word 'geology' stamped on the *Habilitation* thesis she started last year, the doors to professorships in ecology or biology departments will remain closed to her.

Over the next 10 years, Bulmahn aims to replace the *Habilitation* with a system of six-year junior professorships. The aim is to give young scientists a chance to set up their own independent research groups at an earlier age, while removing the restrictions imposed by the traditional system. Some 180 million euros (US\$160 million) has been allocated to support the first 3,000 junior professorships, which will be awarded over the coming three or more years. More than 50 of Germany's 300 or so universities have already begun to advertise the new positions.

Bulmahn's plans are the latest in a series of initiatives that aim to give young researchers more independence. Building on a model operated by the Max Planck Society since 1969, the private Volkswagen Foundation and the DFG, Germany's main grant-giving agency, have since the mid-1990s supported junior research groups, in which



Caught in a trap: Madelaine Böhme bemoans the fossilized academic system in Germany.

▶ young researchers are placed in charge of a group typically consisting of a postdoc, a PhD student and a technician. Dozens of these groups are now in existence.

This idea has been taken further at the European Neuroscience Institute (ENI) in Göttingen, a joint initiative between the University of Göttingen and the Max Planck Society that comprises six independent neurobiological research groups and which opened last June. The group leaders are full of enthusiasm. "A junior group position is the



Frauke Melchior believes postdocs need more time.

best you can get, and having several groups working together on several aspects of neuroscience with different equipment means that we really learn to collaborate," says Stephan Sigrist, who leads the ENI's neuroplasticity group. Young researchers are broadly supportive of Bulmahn's new initiative, but many are concerned about the rigid time-frame imposed by the system. Researchers will be allowed to spend no more than 12 years on fixed-term contracts — including PhDs, postdoctoral positions and the new junior professorships, but excluding time spent abroad. Many of those who attended *Nature's* round table argued that these rules could deny many scientists the extended period of postdoctoral training that can be vital in building the skills needed to become a successful lab head. It may also reinforce a trend that in the 1990s saw 15% of German PhD graduates cross the Atlantic.

"The transition from PhD research to full independence is very hard, especially in the life sciences," Frauke Melchior, a biochemist who leads a junior research group at the Max Planck Institute for Biochemistry in Munich, told the round table. "Solid postdoctoral training is not only extremely important to become acquainted with scientific techniques; it is also essential for acquiring the leadership skills you need when heading a group involving several PhD students and technicians." Before she returned to Germany, Melchior spent five years as a postdoc at the Scripps Research Institute in La Jolla, California.

The *Habilitation*, which typically takes five years to complete, at least provides an extended period of supervision and training, argue some senior scientists. At the round table, Thomas Bein, a professor of physical chemistry at the Ludwig-Maximilian University of Munich, called for a twin-track solution, preserving the *Habilitation*. "Helping young scientists get started in a well-established environment and under careful supervision by an experienced professor is not the worst idea," he argued.

For the time being, the *Habilitation* will operate alongside the new junior professorships. Bulmahn's plan calls for it to be phased



Teamwork: the newly opened European Neuroscience Institute offers an interdisciplinary approach.

out by January 2010. But that may be challenged in Germany's constitutional court by opponents led by Hans Zehetmair, the science minister in the Bavarian state government.

Most attendees at *Nature's* round table said they will not mourn the *Habilitation's* demise. Indra Willms-Hoff, who oversees funding of the natural sciences by the Volkswagen Foundation, hopes that the end of the *Habilitation* will open top positions to people with alternative qualifications, such as industrial experience. Junior professorships are not the only route to the academic summit, she told the round table.

Disagreement over terms

Nevertheless, many young scientists are concerned about aspects of the new system. Some fear that it pays little attention to the needs of scientists who want to continue working at the bench, rather than aspiring to leadership roles. "If I could, I would prefer to get a good permanent job in the second rank," Monika Kortenjann, a biologist at the Max Planck Institute for Immunobiology in Freiburg, told *Nature's* round table.

Ralf Baumeister, a professor of genetics at the Ludwig-Maximilian University of Munich, added that the limit being imposed on the total time for which scientists can be employed on fixed-term contracts could be damaging for established research groups. "People who can reliably operate wickedly expensive equipment are basically the guarantors of continuity, and we should not risk losing them to industry," he said.

But the biggest concern about Bulmahn's reform plan is that it may represent an uncomfortable halfway house between the traditional German system and a US-style academic career structure.

Giovanni Galizia, a 38-year-old neurobiologist at the Free University of Berlin, is science policy spokesman for the 'Young Academy', a division of the Berlin-Brandenburg Academy of Sciences and Humanities that promotes the careers of young scientists. He spoke for all attendees at *Nature's* round table when he argued that the junior professorships should be US-style tenure-track

positions — that is, each post should be a stepping-stone towards a full professorship at the same university.

This would mark a big departure from the traditional German system, where scientists who gain a *Habilitation* at one university must seek a full professorship elsewhere. Bulmahn has not stipulated such rules for the new junior professorships, although the new system does demand that researchers move institutions at least once in their career before becoming a full professor.

Some universities are now advertising the new posts as tenure-track positions. But most German states, which finance the universities, seem unlikely to follow this lead. "We are not going to copy the US system," says Bavaria's Zehetmair. "Junior professorships must not automatically lead to a permanent position."

Other researchers are concerned that the new posts are not adequately resourced. "Junior professorships are a half-hearted attempt to create US-like career structures at a low price," Bein told *Nature's* round table. "Junior professors will get far less equipment money and, in the absence of tenure track, most likely considerably less support from their universities than US assistant professors."

Those views are echoed by Helmut Strey, a 38-year-old German biophysicist now working in the United States. Junior professors will be given between 30,000 and 60,000 euros over six years to pay for equipment. That contrasts with the US\$400,000 that Strey received for equipment in 1999 when he started his assistant professorship in polymer research at the University of Massachusetts in Amherst. "From the first day, I was treated as an equal, and my department supports my work 100%," says Strey. "In Germany, young scientists are often regarded as cheap labour."

If Bulmahn's reforms are to succeed, they will not only have to release researchers like Böhme from the straightjacket of the *Habilitation*, but also convince the majority of young German scientists that they truly are valued as independent members of the academic community. ■

Quirin Schiermeier is *Nature's* German correspondent.

Postdocs face hardship across mainland Europe

Those who follow the funding and work abroad may not find a job when they go home.

Sir—As a French postdoc currently living in England, I would like to comment on your News Feature on this topic, “Young, gifted... and spurned” (*Nature* 414, 145; 2001), and the two Correspondences arising from it (P.-L. Chau *Nature* 414, 582; 2001 and M. Bailly, *Nature* 415, 13; 2002).

Comparing France with the United Kingdom in the way that Chau does is inappropriate; and contrary to the impression that may have been given by your News Feature, the situation of French postdocs is (unfortunately) all too common across mainland Europe.

First, Chau’s statement “There are more postdocs for each tenured post in the British system than in the French system” cannot be taken, as he intends, to mean that competition is fiercer in Britain, because it ignores the fact that many postdocs currently living there are from other European countries and wish eventually to find a job in their home country.

I agree that it is easier for a French postdoc to find a permanent position in the United Kingdom than for a UK or other foreign postdoc to find one in France (see the Correspondence by A. Munk *Nature* 413, 772; 2001), but this applies mainly to teaching positions.

Because English is the standard scientific language, it seems unfair to compare France to the United Kingdom by using the ratio of postdocs to tenured staff in the two countries. There are more opportunities for funding postdoctoral positions in the United Kingdom than in France. Even the European TMR (training and mobility of researchers) grants, designed to favour movement of scientists across Europe, are mostly allocated to positions in the United Kingdom: approximately twice as many as for France and 10 times the number for Greece.

Nevertheless, the French system is improving as positions become available to non-French scientists, and I would not call French postdocs “spurned”.

Second, my Spanish, Portuguese and Italian postdoc friends all feel the same as I do. Many PhD students face the choice of either leaving their home country to acquire new scientific experiences but risk being forgotten, or staying at home to build scientific relationships. The price for the latter choice, as Bailly aptly puts it, is to “spend ... time begging for short-term funding which ... barely supports a decent living”.

Most postdocs I know (mostly from

Spain, Portugal or Italy) feel that they made the right scientific choice by working abroad. But they also feel that jobs are given too often to those who stayed, as discussed in your News Feature.

The case of Claire Amadou described in this feature is unfortunately all too common. But she could equally have been called Clara Rodriguez, Vieira or Romanelli.

PhD students also need to make their voices heard

Sir—*Nature* pays a lot of attention to the plight of postdocs everywhere (see for example *Nature* 407, 119 and 408, 637; 2000; and 410, 137; 2001) and North American PhD students (see *Nature* 409, 442; 2001) but only occasionally to European PhD students (see *Nature* 410, 299; 2001). European science needs a sense of European identity before it can speak with one voice.

For this reason, students and young scientists from 15 countries have collaborated to found Eurodoc, the Council for European doctoral and postdoctoral students (www.precarios.org/eurodoc/index.htm). It includes the national PhD, or PhD and postdoctoral, students’ unions of Finland, France, Germany, Italy, Hungary, the Netherlands, Spain, Sweden and the United Kingdom; and local unions from Belgium (Leuven), Estonia (Tartu), Greece (Athens), Ireland, Latvia and Slovakia. Eurodoc’s second general assembly will be in Girona, Spain, from 31 January to 3 February 2002.

The organization is not only revealing similarities in the difficulties that students face throughout Europe, but is also highlighting the differences between education systems. One of Eurodoc’s functions is to identify these disparities and bring them to the attention of the European Union, European governments’ education ministries, scientific societies and the public.

Senior scientists and politicians are already aware of some problems faced by PhD students, but others are unknown or their importance is not properly understood. Eurodoc will therefore explain national systems and report problems, from the student perspective, to the scientific community to encourage international discussion about these European issues.

It is surely time to analyse the situation at the European level (see *Nature* 413, 768–770; 2001). What are the odds in favour of young scientists in different European states building a scientific career in their home countries?

Lounès Chikhi

Department of Biology, Darwin Building,
University College London,
London WC1E 6BT, UK

T. L. Raoul Tan

Former president, Eurodoc, and Department of Cell
Biology and Genetics, Erasmus University,
PO Box 1738, 3000 DR Rotterdam, Holland

Cultural heritage shared

Sir—Your News report about the law covering repatriation of Native American items in California (*Nature* 414, 139; 2001) was even-handed and correct—it can be difficult to follow the NAGPRA statute.

As a retired director of the Harvard Peabody Museum, I know the Tlingit totem pole shown in your photograph and it has been returned to the Tlingits. It was a gift from a Mr Harriman, who obtained it from a professional ‘collector’ who took four such poles in around 1900 from a Tlingit village that was presumed to be deserted. We now know the Tlingits had moved away only temporarily, to protect themselves from a smallpox epidemic.

We have had very good interactions with that tribe. In the mid-1970s (while I was director), Rosita Worl, a Tlingit student in our department of anthropology, helped me put together a major exhibition using the museum’s great early (1860s) collection of Tlingit artefacts. Through Ms Worl, I was able to get help from people in Alaska to curate the exhibition; I brought Nathan Jackson, a Tlingit, to the museum for about five months to make a large wall carving. He, Ms Worl and about 10 more Tlingits did a series of public dance performances.

When the museum received the request for the repatriation of the pole, the current director, Rubie Watson, asked Nathan to make a new pole for the museum. On 19 November, Jackson and his family, Dr Worl (who got her PhD in anthropology from Harvard recently) and other Tlingits danced again in the museum under the shadow of the new pole.

Stephen Williams

PO Box 22354, Santa Fe, New Mexico 87502, USA

Macromolecular marvels

Landmarks along the highway leading to an understanding of proteins.

Nature's Robots: A History of Proteins

by Charles Tanford & Jacqueline Reynolds
Oxford University Press: 2001. 312 pp.
£18.99, \$27.50

Henryk Eisenberg

The story of proteins — the robots of this admirable book's title that do everything from recording visual images and flexing muscles to providing us with epicurean pleasures and better laundry detergents — begins early in the nineteenth century. The name 'protein', which means 'standing in front' or 'in the lead' in Greek, was suggested by Jöns Jacob Berzelius in an 1838 letter to Gerrit Mulder, who had analysed the carbon, hydrogen, oxygen, nitrogen, phosphorus and sulphur content of egg albumin and calculated its formula. *Eiweiss*, or egg white, is, in fact, the archetypal name for protein and is still in the German scientific vocabulary today. Mulder published his results in *Annalen der Pharmacie*, but got into a serious argument with its editor, Justus von Liebig, the leading chemist of the day. Mulder's formula for the proteinaceous *Grundstoff* (core substance), based on the content of a single atom of sulphur and one of phosphorus in proportion to a large number of hydrogen, oxygen, nitrogen and carbon atoms, proved correct and should have established the macromolecular nature of the substance.

A little later, in 1840, Felix Hoppe-Seyler prepared the first crystals of haemoglobin, and from the iron content in the prosthetic group (haemin) associated with the protein (globin), he inferred what also turned out to be the correct subunit molecular weight of the protein. But a debate about whether proteins were macromolecules — as opposed to colloidal aggregates — and, indeed, whether macromolecules could exist at all, raged for the best part of a century. In one of their most illuminating chapters, Charles Tanford and Jacqueline Reynolds trace the history of the macromolecule idea, and show that biochemists were, conceptually, far ahead of their chemist contemporaries, who needed Hermann Staudinger's work on synthetic polymers in the 1930s to be persuaded.

The authors vividly bring to life the triumphs of the great pioneers, who laid the foundations of a subject that today stands at the pinnacle of biological and medical science. Emil Fischer discovered the peptide bond and succeeded in synthesizing short peptides, and Gilbert Adair made exquisitely precise osmotic-pressure measurements to reveal the tetrameric nature of the haemoglobin molecule. Theodor Svedberg



Protein pioneers: (clockwise from top left) Gerrit Mulder, Theodor Svedberg, Dorothy Wrinch and, together, Irving Langmuir (left), J. D. Bernal and Dorothy Hodgkin.

invented the analytical ultracentrifuge, with which he confirmed Adair's results and generated so much information on the physical nature of proteins. Then came the discovery of how electrostatic charge functions in determining the properties of proteins, followed by the introduction of electrophoresis and chromatography for separating the constituents in mixtures. The first studies on denaturation were followed, in 1951, by Linus Pauling's model-building, which identified the α -helix and the β -sheet as the fundamental structural building-blocks of proteins. There were also wrong turns on the highway of progress, such as the absurd cyclol hypothesis of the mathematician and amateur chemist Dorothy Wrinch, unaccountably espoused by the great Irving Langmuir — the story is entertainingly retold by Tanford and Reynolds.

The hydrophobic effect, a key to understanding the thermodynamics of proteins — associated in the second half of the twentieth century with the names of Walter Kauzmann and the authors themselves — turns out to have an older lineage, going back to Langmuir and G. S. Hartley some decades earlier. But the critical work that demonstrated the unique link between amino-acid sequence and structure, and the principle that a protein chain can fold itself unassisted into its native conformation when it is synthesized,

is strangely neglected here: the name of Christian Anfinsen, who won the Nobel prize for these achievements, appears only in passing. It was Anfinsen and his co-workers who showed in 1957 that an enzyme, ribonuclease, could be fully unfolded, with its disulphide bonds broken, and then made to recover its structure and activity. Incorrectly paired thiol groups could be 'unscrambled' and the correct pairing regained. And finally, Anfinsen and colleagues synthesized ribonuclease *ab initio* from its amino acids and showed that the product had full enzymatic activity — the ultimate proof of the proposition.

Tanford and Reynolds very fittingly celebrate the achievements of the crystallographers Max Perutz and John Kendrew and their colleagues, whose heroic efforts over several decades made manifest the structure of proteins and revealed before our eyes the principles of their design — previously only matters of inference and conjecture. This leads us to the modern age of protein expression, site-directed mutagenesis and designer proteins.

Tanford and Reynolds are sometimes a little hard on those who got things wrong. It was Anfinsen again who said that he and some distinguished collaborators had made a bad mistake in their early work on the structural basis of ribonuclease activity: it

took him, he wryly observed, 15 years of toil to correct the error. Aside from a few such matters of balance, I would have liked to have seen more explanatory diagrams, and I also suspect that some parts of the book may be difficult for the outsider to follow. But anyone interested in proteins will find *Nature's Robots* an absorbing and often exciting story, as well as a major contribution to scholarship.

Henryk Eisenberg is in the Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel.

Light on the making of psychiatry

Undertaker of the Mind: John Monro and Mad-Doctoring in Eighteenth-Century England
by Jonathan Andrews & Andrew Scull
University of California Press: 2001. 386 pp.
\$35, £24.95

Akihiro Suzuki

Psychiatry in the eighteenth century used to be regarded as the heart of darkness. The age of holy madness, as seen by medieval Christendom, and that of wise fools in Renaissance theatre, had gone, and the humanitarian reform of the nineteenth century was yet to come. Historians of various, often opposite, inclinations condemned eighteenth-century madhouses and mad-doctors (as psychiatrists were then called). Only with Roy Porter's ground-breaking book *Mind-forg'd Manacles* (Athlone, 1987) did historians begin to see psychiatry in the

eighteenth century in a different light. Since then, several key publications, the most notable of which is R. A. Houston's *Madness and Society in Eighteenth-Century Scotland* (Oxford University Press, 2000), have shown that Porter's emphasis on diversity and dynamism in eighteenth-century psychiatry is essentially correct. Psychiatric modernity, with both its positive and negative aspects, was established — or at least was being gestated — in the period of the Enlightenment. On the positive side, humane attitudes towards the patients and psychological understanding of mental diseases were being formulated in the upper and middle classes. On the negative side, the threat posed by psychiatry to the principle of individual liberty was clearly recognized, thanks to some well-publicized scandals. Both the triumph and shame of psychiatric modernity were present in eighteenth-century Britain.

The major focus of this fresh contribution to the making of English psychiatry is John Monro, the second of the Monro family who served as physicians to the Bethlem Royal Hospital in London (or 'Bedlam', as it was known) for more than 120 years. Particular odium used to hang over Bethlem and its doctors in the eighteenth century — the hospital's practice of allowing curious visitors to watch its inmates was discontinued only in the 1770s. In 1815–16 it was revealed by a parliamentary select committee that a man called James Norris had been kept in chains there for more than 10 years. He became a poignant symbol of neglect, and Bethlem a byword for the bad old, unreformed psychiatry of the Georgian era.

Jonathan Andrews and Andrew Scull have joined forces to rescue John Monro from the odium and to set the record

straight. Lacking personal papers or voluminous works by their protagonist, they turned to several well-recorded incidents in which Monro was involved in his professional capacity. Intellectual aspects of psychiatric theory, especially the question of the definition of madness, are analysed. The authors give a fine revisionist account of the so-called Battie–Monro debate of 1758, in which Monro answered criticisms levelled at his father and Bethlem Hospital by William Battie, physician to St Luke's Hospital for the Insane, London's newly established, rival institution.

The ever-present pitfall of wrongful confinement is examined using the colourful case of Alexander Cruden — or 'Alexander the Corrector', as he called himself. Cruden believed that God had called him to reform the manners of the age, and he provided one of the earliest criticisms of psychiatric care from the perspective of one who had gone through it. A difficulty of a different sort, namely the delicate business of treating an insane member of the aristocracy at home, is also examined, focusing on the psychiatric case of Lord Orford.

Yet another delicate question of psychiatric testimony in a legal context is investigated in a discussion of two high-profile cases: the execution of Earl Ferrers for murdering his steward, and Margaret Nicholson's attempt to kill King George III. Another chapter investigates the business of keeping private madhouses — the key phenomenon in understanding the growth of psychiatry in Britain in the eighteenth century. If Monro's high profile, prestige and notoriety as the leading psychiatrist in Georgian London was based on the position he held at Bethlem, his economic prosperity was anchored in the network of metropolitan madhouses.

Through the examination of these diverse aspects of Monro's career, Andrews and Scull have produced an in-depth account of the ambiguity surrounding the making of psychiatry in England. This is a product of impressive scholarship worn lightly, and a careful evaluation of the numerous factors that structured the nascent specialist practice of treating the insane. Each case is contextualized in terms of broader historical trends such as secularization, the rise of sentimentalism, the advent of the consumer society, and Anglo-French rivalry. At the same time, the authors never lose sight of the individuality of their protagonist and his patients or clients. The doctor, the patients and their families are not presented as cogs in a wheel of structural changes, but are depicted as responding and contributing to the creation of new cultural attitudes towards mental disease. The complicated interaction of structural forces and individual agency is finely analysed, and related in elegant prose. No fewer than 50 illustrations adorn the text,



Madhouse? The grim reputation of Bethlem Hospital, also known as Bedlam, was not wholly deserved.

each accompanied by detailed descriptions. Together with the works of Porter and Houston, this book should be essential reading for anyone interested in the making of psychiatry. ■

Akihito Suzuki is in the History Programme, School of Economics, Keio University, 4-1-1 Hiyoshi, Yokohama 223-8521, Japan.

The man who talks a lot

My Life in Science

by Sydney Brenner

BioMed Central: 2001. 191 pp. £14.99

Jan A. Witkowski

In April and May of 1994, Sydney Brenner sat down in front of a video camera and talked to Lewis Wolpert, recording enough to fill 15 videotapes. Errol Friedberg and Eleanor Lawrence have taken sections from the transcript of the interviews and woven them, with linking and explanatory passages added by Friedberg, into *My Life in Science*, an 'autobiography' of Brenner's more than 50 years of research.

I opened this book with considerable misgivings. Of all the ways to set words on paper, transcription is the least satisfactory, as anyone who has had to edit a transcript knows. The most polished phrases can be reduced to mere platitudes or worse, while every hesitation, deviation and *non sequitur* is recorded for the continuing embarrassment of the speaker. I am glad to say that in the present case my fears were misplaced; Brenner, Wolpert, Friedberg and Lawrence have produced an informative and entertaining book, redolent with Brenner's voice.

There are many pleasures in the book. It is interesting to learn the inside stories behind some of the classic experiments in molecular genetics. The experiment to confirm the existence of messenger RNA was conceived in a moment of inspiration in Cambridge, but carried out in more prosaic circumstances in a basement at CalTech — François Jacob dropped phosphorus-32 into the water bath and Brenner carried the ultracentrifuge rotor from the cold room to the laboratory as though it were a relic of a saint, in solemn ecclesiastical procession. The paper was written in a form unfamiliar to today's researchers. Brenner and Jacob, in their enthusiasm for Karl Popper's insistence on the falsifiability of scientific hypotheses, began by proposing three models, two of which they set out to disprove.

The book gives many of Brenner's *bon mots*, which hide some serious insights into how biological research should be done. Take, for example, the 'Don't Worry Hypothesis' — if you have a strong idea that



Brenner: famous for his championship of the nematode that can be grown in a Petri dish.

explains something interesting, don't worry too much about what it doesn't explain. Life is complicated, and even the best idea is not going to explain everything. Without the 'Don't Worry Hypothesis', the double helix would have been rejected because no one understood how the two strands of DNA could be unwound during replication. James Watson and Francis Crick put this seemingly insuperable problem to one side, arguing that, as DNA did get replicated, the cell must have ways of unwinding the helix and these would be discovered sooner or later. It is here that Brenner's 'Occam's broom' comes in handy, to sweep the currently difficult facts under the carpet (at least until it becomes impossible to walk on it). 'HAL' is a related concept. Because organisms are complicated, often we do not have the knowledge to make clear-cut predictions. So the thing to do is, 'Have A Look' — just try something and see what happens. Unfortunately, this eminently pragmatic strategy is unlikely to find favour with funding agencies.

Brenner is best known, perhaps, for his championship of the nematode worm *Caenorhabditis elegans* as an experimental animal. He recounts how he searched through a veritable bestiary of weird and wonderful animals — the protozoa *Paramecium*, *Naegleria* and *Hartmanella*; rotifers; and the fly *Sciara*. But although these organisms had interesting biology, they were not amenable to genetic studies. Then he found nematodes, which have an ideal sex life — from a geneticist's perspective — a nervous system small enough to be described completely, and — especially appealing — they can be grown, like bacteria, in Petri dishes. The rest, as they say, is history; more than 3,000 papers have been published on *C. elegans* since Brenner's classic paper on its genetics was published in 1972. One of these, of course, was the complete sequence of the *C. elegans* genome, the first genome to be determined for a multicellular organism.

This is not a scholarly book and it does

not need to be. However, BioMed Central has done the subject and the editors a grave disservice in two regards. First, there are no photographs, even though there must be innumerable pictures of Brenner from every stage of his career; a few of these would have added greatly to the pleasure of reading the book. Second, there is no index, not even one of names.

Brenner tells us that Fred Sanger described him as "the man who talks a lot", and, indeed, Brenner considers talking — producing a "stream of unconsciousness" — to be one of his great skills. We are fortunate that these skills are recorded and it is good to have more of Brenner's words down on paper. Brenner remarked once that young Turks quickly change into old Greeks. This fate has not yet befallen him, and together with his own *Loose Ends* (1997), *My Life in Science* provides us with a view of the life and works of one of modern biology's most innovative, influential and irrepressible characters. ■

Jan A. Witkowski is at the Banbury Center, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724-0534, USA.

An integrated view of *Bacillus*

***Bacillus subtilis* and its Closest Relatives: From Genes to Cells**

edited by Abraham L. Sonenshein,

James A. Hoch & Richard Losick

ASM Press: 2001. 646 pp. \$149.95

Nicola Illing

Bacillus subtilis has been touted as the model organism for Gram-positive bacteria as well as for microbial cell differentiation for many years. This vision — inspired by work in the late 1950s by John Spizizen and Pierre Schaefner — has come to fruition and is embodied in this comprehensive review of all aspects of *B. subtilis* physiology and genomic information. This text is more comprehensive than an update of the previous *Bacillus subtilis* and Other Gram-Positive Bacteria (ASM Press, 1993), by the same group of editors, in which physiological and genetic information of different Gram-positive bacteria was catalogued in detail, but with little integration between the two. It is inevitable that this new volume will also be compared with the other major prokaryotic reference, *Escherichia coli* and *Salmonella*: *Cellular and Molecular Biology*, edited by Frederick Niedhardt (ASM Press, 1996), and covering the biochemistry of Gram-negative bacteria. This new work not only provides a biochemical view from the Gram-positive perspective of the bacterial fence, but is in a different league with respect to the scope of its coverage.

What is particularly inspiring about this

volume is that it is one of the first of a new generation of reference books that successfully integrates information generated by the sequencing of the *B. subtilis* genome with the many facets of microbial physiology — including cell architecture, chromosome replication and division, metabolism and its regulation, macromolecular synthesis, and adaptation and differentiation. The challenges of annotation — and interpreting the volume of information generated by genome sequencing — are addressed in a separate chapter, and will serve as a useful reference for any group embarking on a genome project.

In addition to focusing on the genes that have been assigned function on the basis of comparative sequence information, all chapters include an analysis of homologues identified in different bacterial species, providing valuable insight into the evolution of specific genes. For example, the genes involved in endospore formation in the genera *Bacillus* and *Clostridium* are compared by a genome-wide analysis searching for homologues among their constituent species. Whereas some of the key sporulation genes — including transcription factors and their regulators — are conserved across *Bacillus* and *Clostridium*, which are both endospore-forming genera, other genes are not. The phospho-relay system that activates *spoOA* — the master regulator of sporulation in *B. subtilis* — in response to the metabolic condition of the cell, the state of chromosomal replication and cell density, is absent in *Clostridium*. This analysis has also shown that homologues of many of the sporulation genes, such as *spoIII*, are also found in bacteria that do not form spores, including *Escherichia coli*. These genes are postulated to have evolved a new function in endospore-forming bacteria.

There is much in this reference for researchers working on genes that originated in prokaryotes. I focused on chapters dealing with stress tolerance, and cell signalling through the histidine kinase phospho-relay cascade, and found a wealth of information that can be applied to many biological systems, such as stress tolerance in plants.

Although each chapter has been written by a different group of authors, a common thread and uniformity of style runs through the book. In spite of its size and breadth of scope, this is a very readable book. Explanations are solidly grounded on experimental data and references. The referencing cannot be faulted and includes a combination of classical and current references. *Bacillus subtilis and its Closest Relatives* sets a new high standard for physiology and biochemistry texts, and is an essential reference for researchers, university libraries and both undergraduate and graduate students. ■

Nicola Illing is in the Department of Molecular and Cell Biology, University of Cape Town, Rondebosch 7700, South Africa.

Science in culture

The Kings' comet

Giotto, Halley and the art of observation.

Martin Kemp

According to legend, it was 11 days after Christmas, on Epiphany, that the Kings brought their gifts to Christ. They arrived at precisely the right spot courtesy of a guiding star, whose behaviour differentiated it clearly from those normally visible in the night sky. If it was a star outside the normal run, what did it look like?

For most medieval artists, the conventional pointed star did the job well enough, hovering above the stable in which Christ had been born. But not so for Giotto di Bondone, the great early-fourteenth-century pioneer of naturalism. In his lucid and eloquent image of the *Adoration of the Magi* in the Arena Chapel in Padua, Giotto characterizes the Kings' guiding light as a flaming ball streaking across the heavens. The eight-pointed star at the centre of the array dissolves at its rear into a trailing tail of streaky fire. Less prominent than when first painted, through losing some of its painted gold, it still has a striking effect, unparalleled in earlier representations. It is easy to believe that Giotto had witnessed the advent of a comet. But which one?

By far the best candidate is Halley's comet, which was visible from mid-September to November in 1301. The date works well enough, for Giotto painted his image in the years immediately preceding 1305. The comet's visit was noted in various written records, and one Italian chronicler described it as

having "great rays of smoke behind it".

Giotto's ability to depict the comet so convincingly was due to his innovative naturalism. More than a century before the invention of linear perspective, he had experimentally determined how to create the effect of solid forms, overlapping in space, with simple architectural features slanted back into the picture. One important tool was his use of a consistently directional light, described as coming from a large window on the west wall of the chapel. Not only does Giotto use this light to sculpt the figures, but he also plots its passage on the stable, as the diagonal beams of the roof fall systematically into shade on its far side.

It is in this context that the comet declares itself in the spectator's mind as an observed phenomenon rather than a conventional pictorial symbol. Embedded precociously in Giotto's image is the as-yet unrecognized potential for the use of empirical science in the naturalistic techniques forged by the artists of the fourteenth and fifteenth centuries.

Giotto's portrayal of the appearance of a 'real star' could not be more fitting for Epiphany, given that the Greek *epi* means 'around' or 'over' and *phainein* 'appearance' or 'manifestation'. It was also a happy choice to name the spacecraft launched in 1986 to investigate Halley's comet after the great Italian artist.

Martin Kemp is in the Department of the History of Art, University of Oxford, Oxford OX1 2BE, UK.



BRIDGEMAN ART LIBRARY

Setting standards

Frank Wilczek

Rough conceptions of universality predate science, and even human consciousness. A swimming fish implicitly assumes that the laws of hydrodynamics are universally valid, and the basic properties of its watery environment invariable. Newton's universal law of gravitation defined an ideal model for classical physics, which was consciously emulated in the early study of electricity, magnetism and speculative atomic theories. Spectroscopy revealed the accurate and detailed similarity of terrestrial and celestial matter. But it was only in the twentieth century that universality evolved from a metaphysical assumption into a physical concept, the precision, origin and limits of which could be powerfully addressed.

The deepest and most revolutionary insights arose from quantum theory. They are at two levels. First is the very basic fact that matter is built up from vast numbers of copies of a few fundamental components (such as electrons, quarks, photons and gluons). The properties of these elementary building-blocks are always and everywhere the same — universal. If this were not the case, there could be no laws of chemistry, because every atom would have its own quirky properties.

From the perspective of classical physics, this universality is both non-essential and surprising. If elementary components such as electrons were not quite precisely identical, say if their masses varied over a range of a few

parts per billion, it would be conceivable that future observations, more accurate than those possible today, might reveal small differences between them. Indeed, such differences might be expected to arise, because over its long lifetime each electron might change as a result of accidents of its individual history. And this, of course, begs the question of why they were accurately similar to begin with!

In quantum mechanics, the view is fundamentally shifted, as there is a radical difference between two particles being precisely identical and being merely similar. If *A* and *B* are identical, then when *A* travels from position *x*₁ to *y*₁ and *B* from *x*₂ to *y*₂, the final result is the same as when *A* goes to *y*₂ and *B* to *y*₁. In quantum mechanics, the basic goal is to calculate amplitudes, the square of which gives the probability of an event. To get the total amplitude to find particles at *y*₁, *y*₂, we must add (for identical bosons) or subtract (for identical fermions) the amplitudes for these two possibilities, and then square to get the probability. If the particles are distinguishable, then so are the two final states, so we must square first, then add.

The mathematics of quantum statistics works only if we have precisely identical — that is, indistinguishable — particles. It underlies a host of observed physical phenomena, from lasers to neutron stars, as well as the existence of the periodic table and the recondite details of quark and gluon scattering. Thus, the universality of building-blocks is a rigorously demonstrable experimental truth. But why?

In attempting to reconcile the principles of quantum mechanics with the demands of special relativity, we cannot deal directly with individual particles. To construct relativistic quantum theories we need (quantum) fields. Particles arise as secondary manifestations — excitations of the fields. Thus, all electrons are indistinguishable because each is minted at the same press.

The second level of revolutionary insight is that of construction. Even with identical electrons, classical physics would not arrive at identical atoms, but a continuum of 'solar systems'. In quantum theory, it is different. Put together, the basic components snap into a few definite structures. Energy levels are separated by quantum jumps. If a composite system in its minimum-energy (ground) state is probed with insufficient energy to excite it to the next level, it remains in its ground state. Thus, from universal building-blocks we get universal, reproducible structures and reactions — in a word, chemistry.

This brings us to the powerful idea of 'emergent' universality. A theory of low-energy

Universality

The general occurrence, in widely different circumstances, of a common structure or behaviour.

gy behaviour can ignore structure that is definitely present (and that would be revealed to high-energy probes) yet still be perfectly rigorous. Conversely, one can have many distinct alternative theories of fundamental interactions at high energy, or equivalently short distance, that all map onto exactly the same effective theory for low energies.

For fundamental physics, emergent universality is both a blessing and a curse. It provides a sense in which understanding, once achieved, will never be superseded. But in doing so it marks off wide domains as immune to further discovery. It shows how the reductionist programme, to base physical science on ever more rigorous laws, could end "not with a bang, but with a whimper".

There is another 'miracle' of emergent universality, with the emphasis on 'Universe'. Distant parts of the Universe are broadly similar to one another, yet even physical laws that apply universally can nevertheless engender non-universal behaviour, if they act within different environments. Big Bang cosmology postulates physical conditions that change with time, and different parts of the Universe need not have been accurately synchronized. Also, there can be pervasive fields, the values of which could vary in space. The discovery of at least one such field, the axion, is widely anticipated. The idea of inflation reclaims the universality of the observed Universe, for if it originated through expansion of a tiny patch, there is less scope for variation.

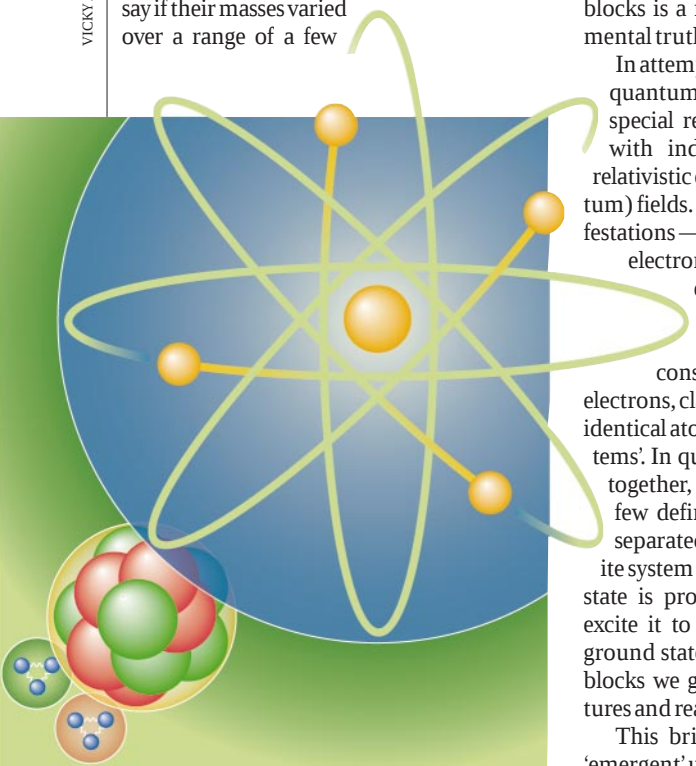
A better understanding of the origins of universality teaches us to conceive its possible limitations. Emergence, in emergent universality, need not be complete. Small energies are not infinitely small, and rare processes can hint at a deeper-lying structure. Inflation does not occur by an infinite factor, and would not enforce perfect uniformity even if it did. Indeed, some (quantum?) fluctuations are required to seed the formation of structure in the Universe — which is to say, its deviation from perfect universality! As always, great answers lead to great questions. ■

Frank Wilczek is at the Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307, USA.

FURTHER READING

Guth, A. *The Inflationary Universe* (Addison Wesley, Boston, 1997).
Weinberg, S. *Dreams of a Final Theory* (Pantheon, New York, 1992).
Wilson, K. *Sci. Am.* **241**, 158 (1979).

VICKY ASKEW



Quantum effects of gravity

Thomas J. Bowles

The effects of gravity and quantum mechanics rarely overlap because of the different scales involved. An experiment with ultracold neutrons has now been able to probe both simultaneously.

The visible effects of gravity usually occur at large scales: gravity governs the path of projectiles, and the motion of stars and planets. By contrast, the effects of quantum mechanics — one of the great successes of twentieth-century physics — are usually only observable at the atomic scale. In the quantum realm, the gravitational force is so weak that it is difficult to observe quantum effects caused by gravity. But on page 297 of this issue, Nesvizhevsky and collaborators¹ report an experiment in which they observed quantum effects of gravity on the behaviour of ultracold neutrons (UCNs). These neutrons have kinetic energies so low that they can be trapped by gravity above a reflecting surface.

Neutrons can be reflected from surfaces when the repulsive force arising from the potential barrier at the surface is greater than the downward velocity component of the neutron perpendicular to the surface. Because the potential barrier is so small, normally only those neutrons arriving at a shallow angle to the surface will be reflected, whereas those that are dropped vertically on to the surface will be absorbed, or transmitted to the other side. But the total velocity of UCNs is so small (less than 8 m s^{-1}) that they are always reflected, regardless of the angle of incidence. Another important characteristic of UCNs is that their gravitational interaction is about as strong as their kinetic energy. A UCN that leaves the surface in the vertical direction is slowed down and eventually turned around by gravity. These two properties allowed Nesvizhevsky *et al.*¹ to construct a trap in which the UCNs are limited in their motion by a reflecting surface from below and by gravity from above.

The trap built by Nesvizhevsky and colleagues can be described as a potential-energy well. A particle in a potential well is trapped because it doesn't have enough energy to escape the well. Classically, a particle inside the well can have any energy as long as it is less than the escape energy. But in quantum mechanics, particles in a potential well are only allowed discrete energy values. In the case of electrons inside an electromagnetic potential well, the discrete (quantum) states are directly responsible for the structure of atoms. The gravitational potential well created by Nesvizhevsky *et al.* also has discrete energy states, the lowest ($n = 1$ state) being at 1.41 peV (1 peV is 10^{-15} electron volts),

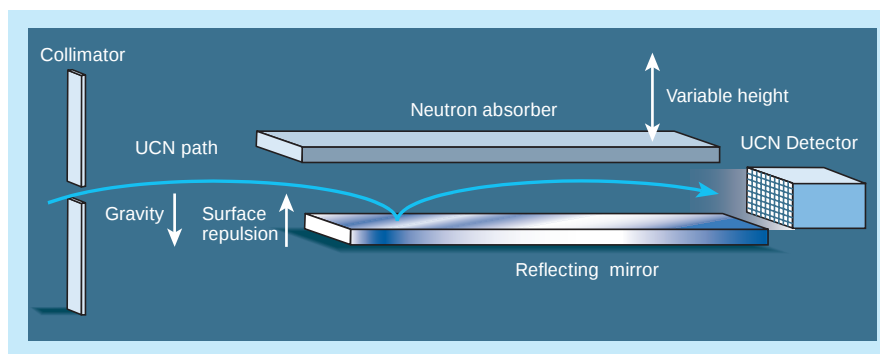


Figure 1 A system for analysing the flight path of ultracold neutrons (UCNs). In their experiment, Nesvizhevsky *et al.*¹ take a beam of UCNs and let them fly horizontally above a reflecting mirror. Providing that all other forces (excluding gravity and repulsion by the mirror) are eliminated, the UCNs will follow parabolic trajectories through the system. According to quantum mechanics, if one measures the vertical velocity component of a UCN in the trap, there are expected to be discrete values corresponding to the quantum energy levels of the trapped neutrons. By adjusting the height of a neutron-absorbing layer above the mirror, Nesvizhevsky *et al.* identify the lowest-energy UCNs transmitted through the system, and therefore the first quantum state of UCNs allowed in the trap. This is a first indication of quantum effects caused by gravity.

corresponding to a vertical UCN velocity of 1.7 cm s^{-1} . A UCN with this velocity can travel only $15 \text{ }\mu\text{m}$ in the vertical direction before it is turned around by gravity.

So in this experiment, quantum mechanics requires that a UCN inside the trap cannot have a vertical velocity component of less than 1.7 cm s^{-1} . It can have higher values, but a UCN with greater energy must have a vertical velocity that corresponds exactly to one of the higher energy states ($n = 2, 3, \dots$). Because the gravitational force acts on the vertical velocity only, there is no potential well in the horizontal direction in the trap, and the neutron's horizontal velocity can have any value.

Nesvizhevsky and colleagues used the intense UCN source at the Institut Laue-Langevin reactor in Grenoble, France, to produce a highly focused beam of UCNs. The authors control the vertical velocity component of UCNs entering the trap by adjusting the height of a neutron-absorbing material above a reflecting surface (the mirror in Fig. 1). This system serves to analyse the vertical velocity component of the UCN in the gravity-mirror trap by measuring the transmission of UCNs through the system as a function of the height of the neutron absorber above the mirror.

The authors found that no UCNs were transmitted at all until the absorber was more than $15 \text{ }\mu\text{m}$ above the mirror. In classi-

cal mechanics, neutrons with any value of vertical velocity could be transmitted, so one would expect to see the number of UCNs increase as the absorber is raised. In the quantum picture, until the vertical velocity component of the UCNs coincides exactly with the energy of the first quantum state, no UCNs can exist in the trap, so no UCN can be transmitted. As the height of the absorber increases further, sudden increases in the number of transmitted UCNs are expected whenever their vertical velocity component matches the higher-energy quantum states.

The data show some hint of stepped increases at the values corresponding to higher energy states, consistent with the existence of these states, but they are not yet conclusive. Nonetheless, the evidence for the existence of the first energy state is convincing and confirms that a quantum effect occurs in the gravitational trap. The difficulty of this measurement should not be underestimated. The researchers are measuring a quantum effect caused by gravity that requires a resolution of 10^{-15} eV . Interactions of the neutrons with other fields would normally obscure such a tiny effect, but the neutron's lack of electric charge and the low kinetic energy of the UCNs make such observations possible.

The authors are planning further studies of gravitationally trapped neutrons. The

energy resolution of the system is defined by the quantum uncertainty principle and is determined by the time the UCNs spend in the trap. If this time can be substantially increased, an energy resolution as high as 10^{-18} eV could be achieved. More intense sources of UCNs (which are under development) are required for such precision, and could lead to new studies of fundamental physics. For example, improved tests of the equivalence principle are needed to investigate the interplay of quantum mechanics and gravity. The equivalence principle says that in a uniform gravitational field all parti-

cles, regardless of their mass or composition, fall with the same acceleration (9.8 m s^{-2} near the surface of the Earth). This means that the inertial and gravitational masses of the neutron have to be equivalent. Until now, such assumptions have been hard to check systematically, but the work of Nesvizhevsky and colleagues could provide physicists with a new probe of the fundamental properties of matter. ■

Thomas J. Bowles is at the Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA.
e-mail: tjb@lanl.gov

1. Nesvizhevsky, V. V. *et al.* *Nature* **415**, 297–299 (2002).

protein kinase C, are likely suspects⁵. By activating this kinase, the lipid molecules seem to reduce the activity of a molecule known as IRS-1, a key component of the insulin signalling pathway.

The latest instalment of this story begins with an apparent paradox. Although obesity is associated with insulin resistance and lipotoxicity, so too is a rare human disorder, lipodystrophy, in which fat tissue is absent. As a result, excess lipid accumulates in tissues such as the liver and skeletal muscle, and patients often suffer from a virtually untreatable resistance to insulin. Genetically engineered experimental animals show the same symptoms⁷. In both obesity and lipodystrophy, then, excess lipid in cell types other than adipocytes is associated with insulin resistance. If the lipotoxicity theory is true, then depletion of this intracellular lipid should improve sensitivity to insulin.

Leptin is produced by adipose cells and 'reports' nutritional information to regulatory centres, in the hypothalamus and elsewhere, to regulate the amount of adipose tissue. Treatment of mice with a genetically engineered form of the hormone indeed reduces the amount of adipose tissue. But it also reduces the levels of intracellular lipid in, for example, skeletal muscle, liver and pancreatic beta cells^{1,8,9}, and — as predicted by theory — improves sensitivity to insulin^{6,7}. Treatment of one strain of lipodystrophic mice with leptin depletes the massive fat deposits in the liver and elsewhere, and corrects the animals' diabetes⁷. In a second lipodystrophic strain, fat-cell transplants from normal mice correct the diabetes, but transplants of fat cells that do not produce leptin (from *ob/ob*

Diabetes

Fat in all the wrong places

Jeffrey Friedman

Obesity and a rare, congenital absence of fat cells are associated with damaging levels of fat in various tissues, and diabetes. Leptin helps to remedy these problems by causing oxidation of fatty acids in mitochondria.

Obesity is often associated with insulin-resistant diabetes, and the health consequences of excessive fat can largely be attributed to this connection¹. The cause of the association is not known. But from studies using the hormone leptin to treat severe insulin-resistant (type II) diabetes in experimental animals, it seems that the likely culprit is fat, not in specialized adipose tissue, but elsewhere. Writing on page 339 of this issue², Minokoshi *et al.* add to this picture by showing that leptin can cause fat to be cleared from skeletal muscle and by describing the cellular mechanism involved.

Diabetes is characterized by increased levels of glucose in the blood stream, which results in organ damage. Blood glucose is controlled by insulin, which is secreted by beta cells in the pancreas. Insulin stimulates the uptake and use of glucose by muscle and fat cells (adipocytes), and reduces its output by the liver, thus lowering glucose levels in the blood. Insulin also stimulates lipid biosynthesis in fat and liver cells.

Obesity is generally associated with resistance to insulin-stimulated glucose transport in muscle and fat, and to insulin-stimulated suppression of glucose production in the liver. When insulin secretion cannot meet the increased demand in obese subjects, a rise in blood glucose levels ensues. These events can, to some extent, be reversed by even modest weight loss, which improves insulin signalling and protects pancreatic beta cells. But why is obesity associated with insulin resistance, and weight loss with improved insulin signalling? Various reasons have been proposed, one of which is that excess lipids, particularly lipids in the 'wrong places', can inhibit insulin signalling and also impair beta-cell function^{3–6}.

According to this 'lipotoxicity hypothesis', insulin resistance develops when excess lipids are deposited in insulin-sensitive cell types other than adipocytes (which are uniquely designed to store fat for use as energy during hard times). The other cell types include those of liver and skeletal muscle. The result of excess lipid is an inhibition of insulin action. The precise identity of the lipid factor responsible is not known, although fatty acyl CoA (a biochemically active fatty acid) and/or diacyl glycerol molecules (one glycerol molecule bound to two fatty acids), acting through a form of

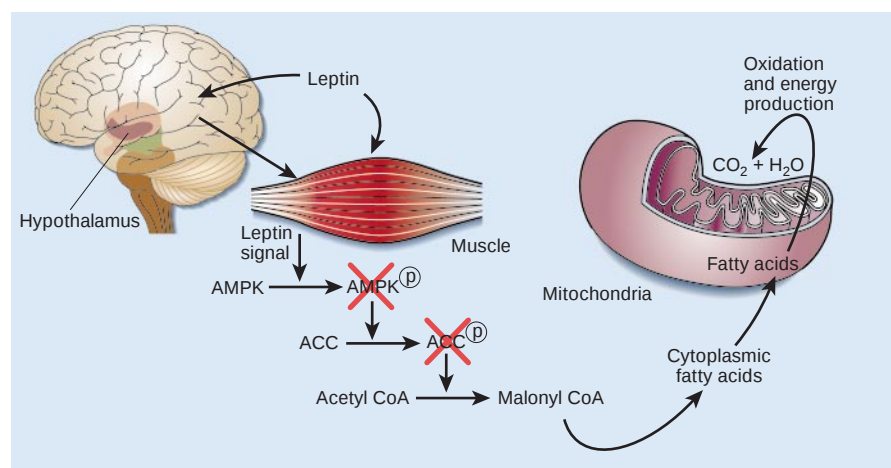


Figure 1 Leptin's control of fat in skeletal muscle². In cells, there is a balance between transport of fatty acids into mitochondria and their subsequent oxidation, and storage of these compounds as triglycerides in the cytoplasm. This balance is regulated mainly by malonyl CoA, a fatty acid that is generated by the enzyme acetyl CoA carboxylase (ACC). Malonyl CoA inhibits transport of fatty acids into mitochondria, thereby preventing their oxidation¹². Leptin causes the phosphorylation of AMP-activated protein kinase (AMPK), which in turn phosphorylates ACC, inactivating it¹³. Leptin thus inhibits malonyl CoA synthesis, leading to greater mitochondrial import and consumption of fatty acids. These events seem to result both from the direct action of leptin on skeletal muscle and from its indirect influence that operates through the hypothalamus.

mice) fail to do so (M. Reitman, personal communication).

Leptin also improves insulin sensitivity in other settings. In normal animals, it reduces cellular lipid content, increases glucose uptake by muscle and increases insulin sensitivity¹⁰. In leptin-deficient *ob/ob* mice, an improvement in the efficacy of insulin is evident at doses that do not affect weight¹¹. In each of these cases, the effects of leptin are evidently not just a consequence of its ability to reduce food intake^{1,7}.

How, then, does leptin clear lipid from non-adipose tissue? This is where Minokoshi *et al.*² come in. They first demonstrate that leptin increases fatty-acid consumption, by oxidation, in skeletal muscle. They also show that leptin activates an enzyme — AMP-activated protein kinase, or AMPK — in skeletal muscle. In cells, a delicate balance controls whether fatty acids are transported into mitochondria and metabolized or are stored in the cytoplasm as triglycerides (Fig. 1). This balance is mainly regulated by malonyl CoA, a fatty acid that is generated by the enzyme acetyl CoA carboxylase (ACC). Malonyl CoA inhibits transport of fatty acids into mitochondria, thereby preventing them from being metabolized¹². AMPK phosphorylates ACC, inactivating it¹³. By activating AMPK in muscle, leptin inhibits malonyl CoA synthesis and shifts the balance towards fatty-acid oxidation and away from fat storage (Fig. 1). These effects are similar to those seen in mice in which the genes that encode ACC have been knocked out¹⁴.

Leptin-mediated phosphorylation of AMPK seems to operate both directly through skeletal muscle and indirectly through the hypothalamus, although the mechanisms involved are not known. We now need to find out whether leptin's lipid-clearing effects on liver and other tissues work in the same way as those on skeletal muscle — interestingly, different forms of the ACC gene are present in skeletal muscle and liver¹⁴.

All in all, the data suggest that we may now have an explanation for the association of obesity, insulin resistance and diabetes. The findings also suggest that leptin may be of benefit not only in obesity but in other settings as well. Some obese subjects are resistant to leptin and do not seem to respond to the genetically engineered form of the protein. However, some obese subjects do lose weight in response to leptin therapy, and the hormone could prove useful in a subset of obese patients, especially diabetic ones¹⁵. Leptin might also be effective in treating human lipodystrophy, a possibility that is now being tested (P. Gorden, personal communication). Lipodystrophy is rare as an inherited disorder, but some HIV-infected patients develop this condition. Indeed, it may be that, in both lean and obese leptin-

sensitive individuals, leptin could reduce excess fat at sites such as the liver and heart, and in beta cells, and prevent the damage it causes.

Finally, there is an evolutionary angle to these results. Leptin's emergence in vertebrates may have helped to prevent excessive weight gain through lipid deposition in adipose tissue in times of surfeit. Excessive weight can easily be imagined to be disadvantageous if, for instance, it makes an animal less able to evade attack. But if lipid in other cells increases the risk of diabetes and other complications of obesity, and leptin acts to reduce this, then this hormone may also have evolved to limit accumulation of lipid in the wrong places⁶.

Jeffrey Friedman is at The Howard Hughes Medical Institute, Rockefeller University, 1230 York Avenue, New York, New York 10021-6399, USA.

Behavioural science

The economics of impatience

Ernst Fehr

In experiments, animals often prefer smaller, immediate rewards over larger rewards that are deferred — thus failing to maximize their total gain. Many people exhibit similar behaviour.

The irresistible cravings of addicts provide an extreme example of short-term behaviour with adverse long-term consequences. The field of study known as experimental and behavioural economics indicates, however, that some of the principles that underpin addictive behaviour are quite common and can help us to understand human behaviour more generally.

At a meeting* held late last year, participants discussed how the tools and concepts of this field can be used to understand, predict and influence people's decisions in circumstances in which some of the rewards or costs of a choice accrue in the future. In trade jargon, these are known as 'inter-temporal choices', and at some time or other we all have to make them. Examples are decisions about savings and the amount of credit-card debt we run up, and about eating habits and even marriage — all of these choices have positive or negative consequences that lie in the future.

If animals have to choose repeatedly between a smaller reward arriving soon (SRS) and a larger reward arriving later (LRL), the SRS is often preferred even though the repeated choice of the LRL would maximize the overall gain¹⁻³. One experiment², for instance, involved hungry pigeons, with an initial trial phase lasting 30 seconds and an outcome phase of 10 seconds. Irrespective

e-mail: friedj@mail.rockefeller.edu

1. Friedman, J. M. *Harvey Lecture* **95**, 107-136 (1999-2000).
2. Minokoshi, Y. *et al. Nature* **415**, 339-343 (2002).
3. Randle, P. J. *Diabetes Metab. Rev.* **14**, 263-283 (1998).
4. McGarry, J. D. *Science* **258**, 766-770 (1992).
5. Shulman, G. I. *J. Clin. Invest.* **106**, 171-176 (2000).
6. Unger, R. H., Zhou, Y.-T. & Orci, L. *Proc. Natl Acad. Sci. USA* **96**, 2327-2332 (1999).
7. Shimomura, I., Hammer, R., Ikemoto, S., Brown, M. & Goldstein, J. *Nature* **401**, 73-76 (1999).
8. Shimabukuro, M. *et al. Proc. Natl Acad. Sci. USA* **94**, 4637-4641 (1997).
9. Shimabukuro, M., Zhou, Y. T., Levi, M. & Unger, R. H. *Proc. Natl Acad. Sci. USA* **95**, 2498-2502 (1998).
10. Kamohara, S., Burcelin, R., Halaas, J. L., Friedman, J. M. & Charron, M. J. *Nature* **389**, 374-377 (1997).
11. Pellemounter, M. A. *et al. Science* **269**, 540-543 (1995).
12. McGarry, J. D., Mannaerts, G. P. & Foster, D. W. *J. Clin. Invest.* **60**, 265-270 (1977).
13. Winder, W. W. & Hardie, D. G. *Am. J. Physiol.* **277**, E1-E10 (1999).
14. Abu-Elheiga, L., Matzuk, M. M., Abo-Hashema, K. A. H. & Wakil, S. J. *Science* **291**, 2613-2616 (2001).
15. Heymsfield, S. *et al. J. Am. Med. Assoc.* **282**, 1568-1575 (1999).

of the pigeons' choices, each trial of the experiment lasted 40 seconds. In each trial, the pigeons could press one of two keys 2 seconds before the end of the initial phase. One key led to 2 seconds of access to a grain hopper right at the start of the outcome phase; the other key led to 4 seconds of grain-hopper access after 4 seconds of the outcome phase. Almost all pigeons preferred the immediate, but more limited, access to the grain. They thus strongly 'discounted' the LRL in favour of the SRS — in other words, they subjectively devalued the LRL relative to the SRS.

The pattern of discounting also gives rise to preference reversals, because animals are very 'impatient' in the short term and relatively 'patient' in the long run¹⁻³. If, for example, the choice between the two keys had to be made after 2 seconds of the initial phase, leaving a 28-second wait until the beginning of the outcome phase, the pigeons preferred the LRL in more than 80% of cases. So if both the SRS and the LRL were shifted into the future, the animals changed their preferences in favour of the LRL, despite the fact that the absolute time difference between the two options remained unchanged. Their behaviour was thus not consistent over time.

One explanation for this is that, throughout evolutionary history, future rewards have been uncertain. An animal foraging for food may be interrupted or, in the case of reproductive opportunities, die before it

*Nobel Symposium on Experimental and Behavioural Economics, Saltsjöbaden, Sweden, 4-6 December 2001.



Figure 1 Ulysses resists the Sirens' call by being bound to the mast of his ship, the ears of his crew having been blocked with beeswax. Saving for the long term presents a similar, but less exciting, dilemma.

is successful. For humans, the promise of future rewards may be broken. And if the risk faced by a person varies over time, he or she applies various discounts to future events^{4,5} and so behaves inconsistently.

But to what extent do people exhibit similar behaviour to that of hungry pigeons? At the conference, evidence was presented that, for many people, the discount rate for short-term events is likewise much higher than for events that have outcomes further in the future (G. Loewenstein, Carnegie Mellon Univ.). For example, when a small reward is due tomorrow, and a larger reward is due one year hence, many people prefer the small reward. But when the small reward is due in one year and one day, and the larger reward in two years, they tend to prefer the larger reward. For events that are far enough in the future, people are prepared to be patient, and so behave inconsistently depending on the time frame. This conclusion constitutes a challenge for economists who, for decades, have assumed that individuals discount future rewards at a constant rate.

This evidence provides an explanation, going beyond a general tendency to discount future events, for why there is conflict between long-term human intentions and our short-term actions. For instance, if their craving is sated, addicts almost universally prefer not to consume substances that are

detrimental to their long-term health. In the craving state, however, they cannot resist the temptation of their favoured substance.

Likewise, when planning for the long term, most people in developed countries intend to save enough for their retirement, eat healthy foods, exercise regularly, watch less television and quit smoking. But such plans require gratification to be delayed, and most of the time most people put off doing anything about their long-term aims even when they know they should act on them⁶. For example, in a 1997 poll in the United States (where social security is much less generous than in Europe), over 75% of those questioned reported that they should be saving more for their retirement. And looking only at respondents who believed they were at an age at which they "should be seriously saving already", the survey found that more than 50% reported being "behind" in their savings and only 6% reported being "ahead".

The divergence between intention and action creates a demand for commitment: people prefer to be bound to their long-term plans in the same way that Ulysses had himself tied to the mast of his ship to prevent himself and his vessel being lured onto the rocks by the Sirens' song (Fig. 1). The positive value of commitment may be one reason why, in many developed countries,

the state forces people to save for retirement through automatic deduction of contributions from their wages. It may also account for why many households hold a large part of their wealth in illiquid forms such as housing or educational qualifications.

It has proved difficult to devise convincing mathematical models of sub-optimal short-term behaviour. One reason for this is that models in which short- and long-term discount rates differ quickly become analytically intractable. Another is that the evidence about the causes of inconsistent and short-sighted behaviour is ambiguous. It could also be that humans simply aren't good at handling choices that involve many different future consequences with different probabilities⁷⁻⁹.

At the conference, however, a new model of the way in which people consume goods and services was presented (D. Laibson, Harvard Univ.). This approach is based on the idea that the divergence of short- and long-term discount rates produces a conflict within the individual, and that the individual is perfectly aware of this. The model is surprisingly successful in predicting certain aspects of economic behaviour, and is much more accurate than a competing model based on constant discount rates. Among other things, the conflict-awareness model correctly predicts the large drop in consumption that typically occurs when people retire, as well as the high rates of credit-card borrowing despite the high interest rates on this kind of debt.

Another group of researchers described how people's saving behaviour can be changed using the insights on time discounting (R. Thaler, Univ. Chicago; S. Benartzi, Univ. California, Los Angeles). People who are impatient in the short term are reluctant to save more for retirement out of their current income because this requires a delay of gratification. But because they are patient in the long term, they are willing to commit themselves to saving more out of future salary increases.

Thaler and Benartzi persuaded a US company to offer its employees the opportunity to commit themselves to saving more out of future salary increases. Most of the several hundred employees were unwilling to save more out of their current incomes. But most were prepared to save more in the future, and over only 28 months the average savings rate rose from 3.5% to 11.8% of income. So the results of basic research on time discounting can be used to help people to implement their long-term plans.

There is, however, more to experimental and behavioural economics than these examples indicate. In recent years, we have gained deeper insights into the patterns of human reciprocity and altruism, into decision-making under conditions of uncertainty and risk, and into the regularities of human



100 YEARS AGO

Mr. G. Archdall Reid contributes to the current number of the *Monthly Review* an instructive and clearly written account of "the rationale of vaccination"... After passing in review the theories which have previously been held to explain acquired immunity, Mr. Reid shows that it is due to an habituation to the toxins of that disease. This result is brought about by the digestion in the blood of the toxins, so that there are present in the animal's blood toxins in all stages of attenuation, from those newly produced by the microbes, and extremely virulent, to those produced in the beginning of the disease and now in a state of great enfeeblement. Up that graduated scale the cells of the animal react till complete immunity is attained. The serum treatment artificially supplies digestive substances and, what is even more important, a scale of attenuated toxins. Applying these principles to the case of small-pox, the necessity for periodical vaccination is established. It is pointed out that, since small-pox is an airborne disease, isolation, by itself, has no greater power of controlling small-pox than the historic old lady with a broom had of sweeping back the Atlantic.

From *Nature* 16 January 1902.

50 YEARS AGO

Patterns of Marriage. It is possible to give only a few of the results of this research. One point that soon became obvious was that like tended to marry like — the intelligent man, the intelligent woman; the neurotic, the neurotic. Sexual attraction played only a minor part in drawing two people together. Among the psychiatrically sound couples 45 per cent claimed to be happily married, 36 per cent considered their marriage satisfactory, 10 per cent unsatisfactory and 9 per cent admitted to being positively unhappy. In the neurotic group happiness or satisfaction in marriage was less frequent. It is clear also from the answers given that children ranked highest in bringing about marital happiness and that other factors, in diminishing order of importance, were as follows: "clerical rating of personality, economic status, intelligence, orgasm adequacy of the female, pre-marital chastity, good looks, stature, rating of personality by test, similarities between husband and wife and test responses". Frequency of intercourse and youth bore no relationship at all to marital happiness.

From *Nature* 19 January 1952.

learning when people with partly conflicting interests interact with each other. For researchers in this field, these are exciting times, not least because we are witnessing a partial reunification of psychology and economics.

Ernst Fehr is at the Institute for Empirical Research in Economics, University of Zürich, Blümlisalpstrasse 10, CH-8006 Zürich, Switzerland.

e-mail: efehr@iew.unizh.ch

1. Rachlin, H. & Green, L. *J. Exp. Anal. Behav.* **17**, 15–22 (1972).
2. Green, L., Fisher, E. B., Perlow, S. & Sherman, L. *Behav. Anal. Lett.* **1**, 43–51 (1981).
3. Bateson, M. & Kacelnik, A. *Behav. Ecol.* **7**, 341–352 (1996).
4. Sozou, P. *Proc. R. Soc. Lond. B* **265**, 2015–2020 (1998).
5. Azfar, O. *J. Econ. Behav. Org.* **38**, 245–252 (1999).
6. Frederik, S., Loewenstein, G. & O'Donoghue, T. *J. Econ. Lit.* (in the press).
7. O'Donoghue, T. & Rabin, M. *Am. Econ. Rev.* **89**, 103–124 (1999).
8. Rubinstein, A. *Eur. Econ. Rev.* **45**, 615–628 (2001).
9. Fehr, E. & Zych, P. *Scand. J. Econ.* **100**, 643–662 (1998).

AIDS vaccines

One step forwards, one step back

Jeffrey D. Lifson and Malcolm A. Martin

New AIDS-virus vaccines induce cellular responses that can contain, but not prevent, infection. Mutations can allow the virus to escape this immune control, emphasizing the challenges in developing an effective vaccine.

As someone accustomed to persevering on a long-term project in which repeated periods of hard work lead to modest progress, only to be followed by setbacks, Sisyphus would be well suited to a career in AIDS-vaccine research. Papers on pages 331 and 335 of this issue^{1,2} illustrate the point. In the first paper, Shiver and colleagues¹ describe how they immunized groups of rhesus macaques against a protein from a monkey AIDS virus, the simian immunodeficiency virus (SIV), effectively stimulating antiviral cellular immune responses. The immunizations did not prevent, but did help to control, subsequent infection with a related virus. But Barouch and colleagues² show that AIDS viruses can mutate to evade such vaccine-induced, virus-controlling cellular immunity, calling into question a vaccination strategy based solely on such responses.

The aim of most vaccines that offer protection against viruses is to stimulate antibody molecules that can neutralize the virus or otherwise help clear the infection, and cellular immune responses, particularly by cytotoxic T lymphocytes (CTLs) that bear the surface marker CD8 and can kill virus-infected cells. But for the human AIDS virus HIV-1 it has proved difficult to generate vaccines capable of inducing antibody responses that neutralize the broad range of virus strains found in patients. CTL responses may be able to cope with a wider range of virus strains, and appear to help control HIV-1 in infected people, so much current attention in AIDS-vaccine research is focused on vaccines that stimulate CTL responses.

Pursuing this approach, Shiver *et al.*¹ systematically compared vaccination strategies using DNA molecules or engineered non-SIV viruses (a vaccinia virus or an adenovirus known as Ad5) to express a single SIV protein, Gag. After immunizing macaques with these potential vaccines, the authors

detected strong and sustained responses by T cells that express CD8, most notably in animals immunized with the Ad5 vaccine — either alone or after 'priming' immunization with DNA.

To assess the efficacy of the vaccines, the authors¹ then challenged the macaques with a highly pathogenic chimaeric simian-human immunodeficiency virus (SHIV), SHIV 89.6P (ref. 3). This virus contains a gene encoding the viral glycoproteins from the HIV-1 outer 'envelope', as well as several SIV genes, including a Gag-encoding gene matching that used in the vaccines. Like HIV-1 and SIV, SHIV 89.6P infects and kills 'helper' T lymphocytes bearing the cell-surface marker CD4 (ref. 3).

Shiver *et al.* found that their vaccination strategies did not prevent infection with SHIV 89.6P, but did modulate its course, most strikingly in animals immunized with the vaccine based on Ad5. Early on, peak levels of virus in the macaques' blood were only slightly lower than in controls. But by 70 days after challenge, vaccinated animals showed markedly lower levels of virus and preservation of CD4-expressing T lymphocytes. This 'partial protection' is similar to that obtained with other vaccine approaches^{4,5} against challenge with SHIV 89.6P. Used clinically, a vaccine that controlled HIV infection might result in a longer period of infection without symptoms, and a decreased risk of transmission.

However, Barouch *et al.*'s work² provides a cautionary counterpoint to these encouraging results, and raises questions about the strategy of controlling infection by using vaccines that stimulate CTLs alone. These authors describe the course of infection in a rhesus macaque that was at first partially protected from SHIV 89.6P by a DNA-based immunization approach⁵. This vaccinated animal showed high peak levels of virus after challenge, but viral levels eventually

declined to below the limit of detection, with no depletion of CD4-expressing T cells⁵. However, during a year of follow-up², virus replication increased to measurable, albeit modest, levels, in association with the emergence of mutations that allowed the virus to elude a dominant CTL response. An acute loss of CD4-expressing T cells and progressive disease ensued.

This is perhaps more disappointing than surprising. Such escape mutations have been detected before in SIV-infected animals⁶, and there is also evidence of HIV-1 evolution in the face of 'effective' anti-retroviral therapy⁷. So we might have expected, in the setting of similarly 'controlled' viral replication in a vaccinated animal, the emergence of mutations that allow the evasion of immune responses. But this nonetheless represents a fundamental and potentially ominous challenge to the idea of containing, rather than preventing, AIDS-virus infections. The frequency and kinetics of such escape will affect the ultimate usefulness of vaccines that are based on this approach.

Moreover, interpretation of the studies^{1,2} is complicated by the fact that the SHIV 89.6P challenge virus produces a course of infection that is very different from that typical of HIV-1 and SIV. Infection with HIV-1 and most pathogenic SIVs is generally characterized by the gradual, progressive destruction of CD4-expressing T cells. In contrast, infection with pathogenic SHIVs such as SHIV 89.6P typically results in the rapid, systemic and irreversible destruction of nearly all of these cells³.

SHIVs that include the HIV-1 envelope gene were originally developed to evaluate how well monkeys are protected by vaccines that target the HIV-1 envelope glycoproteins⁸. But pathogenic SHIVs, particularly SHIV 89.6P, have been more widely used of late, even for studies in which the HIV-1 envelope glycoprotein is not a vaccine component. Although the vaccines have not completely blocked infection, they have generally lowered the post-peak viral load and prevented the rapid, almost complete destruction of CD4-expressing T cells that is characteristic of SHIV 89.6P infection^{4,5}. This alteration of such a dramatic consequence of infection has been interpreted as evidence of an important advance in the development of an effective HIV-1 vaccine.

However, accumulating data indicate that it is comparatively easy to prevent the loss of CD4-expressing cells and to bring about the sustained control of infection through partial inhibition of SHIV replication early in an infection, whether by vaccination^{4,5,9}, passive immunotherapy¹⁰, short-term drug treatment¹¹, or limiting the amount of challenge virus used¹². Moreover, although rigorous head-to-head comparative vaccine studies have not yet been reported, the results obtained with patho-

genic SHIV challenges contrast markedly with those obtained when pathogenic SIV strains are used to challenge similarly vaccinated monkeys. In such studies, sustained control of SIV infection and its consequences has rarely been achieved^{9,13,14}. So we should be cautious when using alteration of the unusual course of disease caused by pathogenic SHIVs, rather than prevention of infection, as a major test of vaccine efficacy.

All in all, the two studies^{1,2} provide both an encouraging step along the road towards an effective AIDS vaccine and a sobering reminder of how long and difficult the road is likely to be. They do provide some direction along the road, suggesting that successful vaccines will need to induce immune responses that are both durable and broad, recruiting both antibodies and cells¹⁵. Unfortunately for Sisyphean AIDS-vaccine researchers, it is likely that the signposts on this road will continue to indicate an uphill grade for the foreseeable future. ■

Jeffrey D. Lifson is in the AIDS Vaccine Program, SAIC Frederick, National Cancer Institute at Frederick, Frederick, Maryland 21702, USA.

e-mail: lfison@avpaxp1.ncicrf.gov

Malcolm A. Martin is in the Laboratory of Molecular Microbiology, National Institute of Allergy and Infectious Diseases, Bethesda, Maryland 20892, USA.

e-mail: mmartin@niaid.nih.gov

1. Shiver, J. W. *et al.* *Nature* **415**, 331–335 (2002).
2. Barouch, D. H. *et al.* *Nature* **415**, 335–339 (2002).
3. Reimann, K. A. *et al.* *J. Virol.* **70**, 6922–6928 (1996).
4. Amara, R. R. *et al.* *Science* **292**, 69–74 (2000).
5. Barouch, D. H. *et al.* *Science* **290**, 463–465 (2000).
6. Allen, T. M. *et al.* *Nature* **40**, 386–390 (2000).
7. Martinez-Picado, J. *et al.* *Proc. Natl Acad. Sci. USA* **97**, 10948–10953 (2000).
8. Li, J., Lord, C. I., Haseltine, W., Letvin, N. L. & Sodroski, J. *J. Acq. Immune Defic. Syndrome* **5**, 639–646 (1992).
9. Barouch, D. H. *et al.* *J. Virol.* **75**, 5151–5158 (2001).
10. Mascola, J. R. *et al.* *J. Virol.* **73**, 4009–4018 (1999).
11. Igarashi, T. *et al.* *Proc. Natl Acad. Sci. USA* **98**, 658–663 (2001).
12. Endo, Y. *et al.* *J. Virol.* **74**, 6935–6945 (2000).
13. Ourmanov, I. *et al.* *J. Virol.* **74**, 2740–2751 (2000).
14. Wyand, M. S. *et al.* *J. Virol.* **73**, 8356–8363 (1999).
15. Dittmer, U., Brooks, D. M. & Hasenkrug, K. J. *Nature Med.* **5**, 189–193 (1999).

Condensed-matter physics

Eavesdropping on spin talk

Jay Kikkawa

Taking advantage of nuclear and electron 'spin interactions' to store and process information is a long-standing goal. A systematic technique for manipulating spin in a semiconductor provides a low-temperature solution.

The currency of mainstream electronics has always been the movement of electric charge. Building blocks of technology, such as resistors, capacitors and transistors, all rely on the position of charge for their action. But electrons are like tiny magnets with north and south poles, a quantum mechanical property referred to as 'spin', and many scientists believe that harnessing spin could push back the frontiers of information processing¹. This endeavour could be revolutionary because 'spintronic' devices can, in principle, function in a manner that has no charge-based counterpart. One example is the ability of electron spins to exchange magnetic information with nuclei in a solid.

Nuclei have weaker spins than electrons and might behave as stable storage elements because they are better isolated from their environment than electron spins. A fundamental trade-off between information storage times and transfer speeds means that optimizing both in a single material would be difficult. But on page 281 of this issue, Smet *et al.*² circumvent this obstacle and show how complex changes in electronic behaviour can both control and monitor the flow of spin information into and out of the nuclei in a semiconductor.

What lies behind this nuclear–electron crosstalk? Electronic and nuclear spins in a

solid are coupled through the 'hyperfine interaction', which allows a pair of spins to undergo spin-reversal or flip-flop (Fig. 1a, overleaf). This means that an electron can flip its spin by simultaneously flipping a nuclear spin in the other direction, provided that energy is conserved. In a solid material such as a semiconductor, 'information' transfer occurs in the form of a change in net nuclear spin. Both electron and nuclear spins tend to align themselves with an external magnetic field, just as an ordinary magnet would. But nuclear spins are almost always overwhelmed by thermal energy to become randomly orientated, so they contribute virtually no net spin within the material.

The nuclear system can acquire a net spin from the electrons through flip-flop as long as energy is conserved. This requirement can be difficult to satisfy because the energy associated with the orientation of a nucleus is much smaller than that associated with an electron. Accordingly, flip-flop requires a compensatory change in, for example, the electron's kinetic energy (Fig. 1b). Smet *et al.*² chose to study a collection of electrons confined to move in a two-dimensional plane by their host semiconductor and subjected to a perpendicular magnetic field. In the absence of electron–electron interactions and in a strong magnetic field, the electrons in such a system typically cannot

change their kinetic energy appropriately to allow flip-flop to occur (Fig. 1c). This would imply that the nuclear and electronic spins in the system studied by Smet *et al.* would be well isolated.

But electrostatic interactions between electrons lead to exotic states that can interact strongly with nuclear spins. Many different electronic arrangements are available to Smet *et al.*, who can tune the properties of their two-dimensional electrons by adjusting the magnetic field or applying a voltage (thereby changing the density of electrons per unit area). The authors exploit this flexibility to probe the behaviour of electronic spins in two-dimensions and to develop a nuclear magnetometer — an electronic tool for measuring the net nuclear spin.

Smet and colleagues² first tune the density

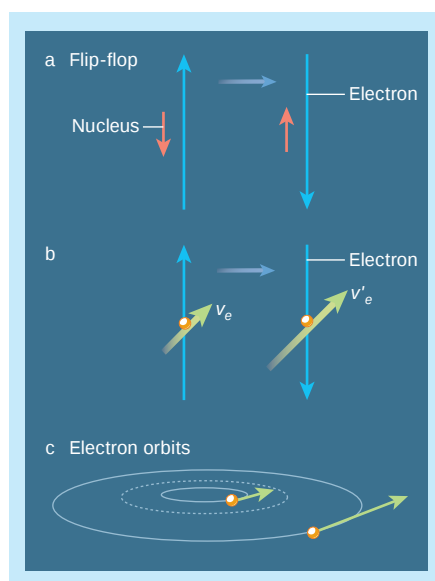


Figure 1 Conservation of energy during spin reversal. a, An electron–nuclear spin pair undergoes spin-reversal (flip-flop). Arrow lengths indicate the magnitude of the energy associated with a spin reversal (the nuclear spin energy has been exaggerated for clarity). b, Nuclear spins require roughly one-thousandth of the energy change of electron spins, so to a good approximation, the nuclear energy can be taken as zero. For flip-flop to work, energy must be conserved, so the change in total electron energy must also be zero. Free electrons can comply with energy conservation by changing electron velocity, v_e (green arrow). Here, the electron lowers its spin energy but increases its kinetic energy by nearly the same amount to allow flip-flop. c, Electrons moving freely in a perpendicular magnetic field follow circular orbits that have discrete sizes dictated by quantum mechanics. So these electrons cannot change kinetic energy by the necessary amount to permit flip-flop because the electron's spin energy does not correspond to the kinetic energy of any of the allowed orbits (solid lines). The orbit with the 'correct' velocity is not available (dotted line), preventing flip-flop.

of electrons in their semiconductor to reduce the energy mismatch between electron and nuclear spin flips, which allows for unusually high coupling between the two spin systems. At the same time, they poise the electronic system at a transition between a configuration with completely aligned electronic spins (maximal net spin) and one with no net spin. The combined result is a net transfer of spin to the nuclear system because when the electronic system changes its net spin it does so, in part, by flipping nuclear spins. In this way, Smet *et al.* show that they can reproducibly program the nuclei to a particular value of net spin.

In addition to allowing nuclear spins to be flipped, the hyperfine interaction also modifies the magnetic field seen by each spin species. So the amount of net nuclear spin has an observable effect on the electronic configuration, which means that the system acts as a sensitive magnetometer. This systematic read/write capability holds tremendous potential for the study of low-energy electronic spin dynamics in two-dimensional electron systems. Smet *et al.* provide a first demonstration by examining how complex electronic arrangements, known as skyrmions, aid flip-flop behaviour (Fig. 2).

For this part of the experiment, Smet *et al.* chose to have the electron spins completely aligned, so the cost for reversing an electron spin far exceeds the nuclear spin-flip energy³. As a result, flip-flop cannot conserve energy and is prohibited. Flip-flop processes are involved in trying to re-establish thermal equilibrium (nearly zero net nuclear spin). So when the authors establish this configuration after imposing a large nuclear moment (as described above), many minutes elapse before the latter disappears⁴. But nuclear relaxation can occur rapidly if skyrmions develop in the electronic system. Skyrmions are circularly symmetrical spin textures involving many electron spins that associate a unit of electron charge with a gradual spin reversal⁵ (Fig. 2). Their presence allows electron spin flips to occur on the nuclear-energy scale, thereby reactivating flip-flop processes⁶. With their nuclear magnetometer the authors monitored the frequency of flip-flop during relaxation, and hence the presence of skyrmions. Their results provide a beautifully detailed confirmation of theory and prior experiments (see ref. 2 and references therein) concerning the response of skyrmions to a magnetic environment.

Smet and colleagues' experiment complements existing techniques for studying nuclear spin dynamics. Nuclear magnetic resonance (NMR) strategies have been successful in probing net electron spin and nuclear spin relaxation in two-dimensional electron systems^{4,7}, and in exploiting non-zero electron spin to detect nuclei residing



Figure 2 The skyrmion regime. Charged spin textures known as skyrmions can have low-energy excitations that enable nuclear–electron flip-flop in conditions when it would not normally occur. The image represents one of many possible skyrmion arrangements. Five skyrmion centres are shown (spins pointing straight up) in a square arrangement. Smet *et al.* have developed a new method that can probe low-energy spin excitations of a skyrmion-filled system. (Figure adapted from ref. 10).

in the same plane as the electrons. New all-optical forms of NMR can stimulate electron dynamics and target nuclei buried within the electron plane directly⁸, but have not yet adapted to specific geometries involving large perpendicular magnetic fields. By contrast, the methods of Smet *et al.* are inherently sensitive to the relevant nuclei and can be performed at the lowest temperatures and highest static magnetic fields available.

Their systematic approach is a recipe for studying nuclear dynamics for all electronic spin states, and supplements advances in nuclear magnetometry by light scattering⁹. This broad range of capabilities is likely to stimulate further musings as to whether nuclear relaxation signatures exist for other electronic phenomena, such as skyrmion crystallization and melting^{6,10} and electron crystallization, which is thought to occur at high magnetic fields and/or low electron density¹¹.

Jay Kikkawa is in the Physics Department, University of Pennsylvania, 209 South 33rd Street, Philadelphia, Pennsylvania 19104-6396, USA. e-mail: kikkawa@physics.upenn.edu

1. Wolf, S. A. *et al.* *Science* **294**, 1488–1495 (2001).
2. Smet, J. H. *et al.* *Nature* **415**, 281–286 (2002).
3. Girvin, S. & MacDonald, A. H. in *Perspectives in Quantum Hall Effects* (eds Das Sarma, S. & Pinczuk, A.) ch. 8 (Wiley, New York, 1997).
4. Tycko, R., Barrett, S. E., Dabbagh, G., Pfeiffer, L. N. & West, K. W. *Science* **268**, 1460–1463 (1995).
5. Sondhi, S. L., Karlhede, A., Kivelson, S. A. & Rezayi, E. H. *Phys. Rev. B* **47**, 16419–16426 (1993).
6. Cote, R. *et al.* *Phys. Rev. Lett.* **78**, 4825–4828 (1997).
7. Melinte, S. *et al.* *Phys. Rev. Lett.* **84**, 354–357 (2000).
8. Salis, G. *et al.* *Phys. Rev. Lett.* **86**, 2677–2680 (2001).
9. Davies, H. D. M., Brockbank, R. L., Ryan, J. F. & Turberfield, A. J. *Physica B* **256–258**, 104–112 (1998).
10. Brey, L., Fertig, H. A., Cote, R. & MacDonald, A. H. *Phys. Rev. Lett.* **75**, 2562–2565 (1995).
11. Wigner, E. P. *Phys. Rev.* **46**, 1002 (1934).

Cell biology

Chloride channels are different

Thomas J. Jentsch

Proteins that conduct chloride ions are vital for a range of cellular processes. The long-awaited crystal structure of a chloride channel shows what these proteins look like, and gives hints about how they work.

Ion channels are proteins with a seemingly simple task — to allow the passive flow of ions across biological membranes. But this process requires more sophistication than one would imagine. Channels must be highly selective for a particular type of ion, yet must also maintain high transport rates and be able to regulate ion flow by switching it on or off (a process called 'gating'). Most attention has been directed towards cation channels, probably because of their role in generating electrical signals in nerve cells. Three years ago Roderick MacKinnon and colleagues revealed the crystal structure of a bacterial potassium channel¹ — a representative of a large family of cation channels. In another spectacular breakthrough, described on page 287 of this issue², the same group now reports the crystal structures of two bacterial chloride channels. They prove to be amazingly different to cation channels.

These new structures² are very welcome, in part because surprisingly little is known about how chloride (Cl^-) channels work. We do know that they come from several gene families³, and that many are gated by voltage (electrical differences across the membrane in which the channel sits). On the basis of elegant biophysical studies by Miller and White⁴, who showed that the electric ray *Torpedo* contains large amounts of a peculiar anion channel, the first voltage-gated Cl^- channel was cloned by my group⁵ in 1990. We named it ClC-0 , as we assumed that it would found a family of Cl^- channels. This turned out to be true: ClC channels are found in all kingdoms of life, with humans alone having nine different ClC genes³.

Some mammalian ClC channels are found in the outer (plasma) cell membrane, whereas others reside in membranes of intracellular organelles. Plasma-membrane ClC channels may stabilize the voltage of excitable cells such as muscle cells, or transport salts across cell layers; for example, mutations in ClC-Kb or its associated subunit⁶ impair salt reabsorption in the kidneys³. Intracellular Cl^- channels may counterbalance the electric current produced by proton pumps, and mutations in ClC-5 , for example, impair the transport of normally acidic vesicles in a kidney disease⁷. The function of bacterial ClCs such as those now crystallized² is not yet clear.

What about the structure and mechanism of the ClC channels? Initial insight came from biophysical analyses of mutated channels; such studies suggested that ClCs consist

of two identical proteins (or subunits), each with its own pore^{8–10}. But, as might be expected, this architecture has made it hard to identify specific amino acids that contribute to the pores — mutations that changed the ion-permeation properties of ClC channels were scattered over much of the amino-acid sequence, and their effects were not very pronounced. (This contrasts with tetrameric K^+ channels, in which four identical segments from each subunit come together to build a single pore.) Another complication was that almost all mutations that changed the properties of the ClC pore also affected gating¹¹.

No wonder, then, that the ion-channel community has been eagerly awaiting a high-resolution crystal structure of a ClC channel. A first step was made last year¹², with a cryo-electron microscopic analysis of two-dimensional crystals of the bacterial EcClC protein at 6.5-Å resolution. The study confirmed that ClC channels are homodimers, and suggested that two areas of lower electron density off the central axis represented two pores.

Dutzler *et al.*² now reveal the three-dimensional crystal structure of two bacterial ClC proteins in unprecedented detail. Their high-resolution (3.0 Å) X-ray analysis allows peptide helices, amino-acid side chains and, amazingly, even the Cl^- ion within the pore to be identified. The crystal structure should convince remaining sceptics¹³ that ClC channels are double-barrelled: two proteins contact each other at a broad interface formed by four helices from each protein. The two pores are not found at this interface, but are completely contained within each subunit, exactly as deduced from the mutational analyses^{8,10}. This contrasts starkly with many other channels, in which four or five identical or struc-

turally related subunits jointly form one pore (Fig. 1). But channels with several pores are not as exotic as they might seem; for instance, aquaporin water channels are quadruple-barrelled. The physiological advantage of having several pores is not clear in either case.

The structure² also helps to explain why previous biochemical analyses of ClC -channel topology produced confusing results. If one looks at the intermingling of helical protein segments in the structure, it's no wonder. These segments often do not cross the entire width of the membrane, and are often severely tilted rather than perpendicular to the membrane (Fig. 2). Each subunit has an internally repeated pattern, not recognized previously, with the two parts having opposite orientations in the membrane (an 'antiparallel' architecture).

At the narrowest part of the pore, there is no positively charged amino-acid side chain that could lead to strong electrostatic binding of Cl^- ions. Otherwise, Cl^- would stay put instead of moving on through the pore. Instead, Cl^- is coordinated by amino acids at the ends of several helices ('D', 'N' and 'R') that, if electrically polarized, would interact with it favourably, but not too strongly. In K^+ channels, the ion is coordinated by amino acids at the ends of four parallel helices coming from the extracellular side. But in the ClC channels² the helices are antiparallel, coming from both sides of the membrane to coordinate Cl^- next to the central plane of the membrane.

Surprisingly, just above the chloride-binding site there is a negatively charged side chain of a glutamate amino acid (Fig. 2), which would be expected to hinder ion transport severely. Dutzler *et al.*² speculate that this side chain acts as a gate, swinging out to open the channel. Conceivably, this movement could be facilitated by electrostatic repulsion when a Cl^- ion enters the pore. This provides an attractive, though speculative, explanation for the observed gating of ClC-0 by anions¹¹.

Many residues that turn out to line the pore² were in fact mutated in earlier studies. For instance, subtle mutations^{8,14} of the key

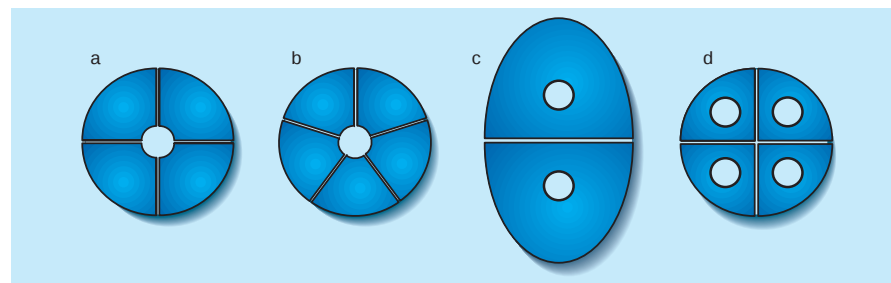


Figure 1 The different ways in which ion channels composed of more than one protein can form pores. **a**, In tetrameric potassium channels, a single pore is formed by four identical or structurally similar proteins (subunits). **b**, In 'ligand-gated' cation or anion channels (such as channels that detect the neurotransmitters acetylcholine or GABA), the single pore is formed by five identical or structurally similar subunits. **c**, Chloride channels from the ClC family are dimers, in which each subunit has its own pore. **d**, Aquaporin water channels are tetramers, again with one pore per subunit.

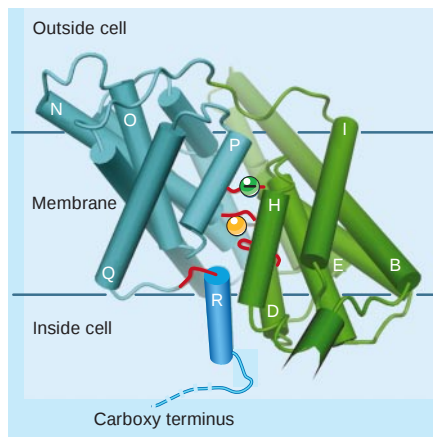


Figure 2 Chloride-channel structure, as revealed by Dutzler *et al.*². This is a side-on view of one of the two subunits of a CLC channel, looking at the side that interacts with the other subunit. α -Helical structures are indicated by cylinders, coloured in green and cyan to indicate the two 'repeated' units. Regions directly involved in formation of the ion-permeation pore are in red. A negative charge that might be involved in gating the channel is shown as a green circle; it is just above where the chloride ion (shown in yellow) is coordinated. The amino terminus of helix R (bright blue) contains a tyrosine amino acid, which helps to coordinate the Cl^- ion; the carboxy terminus emerges into the cell. In mammalian CLCs, helix R is connected to a large intracellular region, indicated by the dashed line. Figure modified from ref. 2.

chloride-coordinating amino acids at the ends of helices D and R changed the channel's ion selectivity, conductance and gating. (But, perhaps surprisingly, the mutant channels were still selective for anions rather than cations.) Intriguingly, the carboxy-terminal end of helix R sticks out into the cell² and, in all CLCs from higher organisms, is connected to a large intracellular segment; some mutations in this segment affect gating drastically¹⁵. So it is tempting to speculate — as Dutzler *et al.*² do — that helix R can transduce intracellular events into channel gating. If so, by pure coincidence, 'R' might stand for 'regulatory'.

Finally, CLC channels — unlike K^+ channels — do not have a large aqueous cavity at one side of the pore. Instead, the CLC pore has a more symmetrical, hourglass shape, whose internal surfaces (except in the narrowest point) have positive charges to attract Cl^- ions. So the crystal structure of CLC channels has revealed many delightful contrasts to cation channels.

What next? The details of the ion-permeation process, including whether or not several ions occupy the pore region simultaneously¹¹, remain to be worked out. And the gating mechanism is difficult to understand from the static snapshots provided by crystals, although the negative charge mentioned above is an intriguing candidate for a gate at a single pore. In addition, in at least two non-

bacterial members of the CLC family, CLC-0 and CLC-1, there is another gate that closes both pores of the dimeric channel at the same time^{4,8,9,16}. This process may require both subunits to change in a way that depends on their extensive intramembrane contacts. Alternatively, intracellular portions of CLC-0 and CLC-1 that are not present in the bacterial channels now crystallized may come into play. Studies of single bacterial CLC channels might be crucial in answering such questions.

Thomas J. Jentsch is at the Centre for Molecular Neurobiology, ZMNH, Hamburg University, Falkenried 94, D-20246 Hamburg, Germany.
e-mail: jentsch@zmnh.uni-hamburg.de

1. Doyle, D. A. *et al.* *Science* **280**, 69–77 (1998).
2. Dutzler, R., Campbell, E. B., Cadene, M., Chait, B. T. & MacKinnon, R. *Nature* **415**, 287–294 (2002).
3. Jentsch, T. J., Friedrich, T., Schriever, A. & Yamada, H. *Pflügers Arch.* **437**, 783–795 (1999).

Arch. **437**, 783–795 (1999).

4. Miller, C. & White, M. M. *Proc. Natl Acad. Sci. USA* **81**, 2772–2775 (1984).
5. Jentsch, T. J., Steinmeyer, K. & Schwarz, G. *Nature* **348**, 510–514 (1990).
6. Estévez, R. *et al.* *Nature* **414**, 558–561 (2001).
7. Piwon, N. *et al.* *Nature* **408**, 369–373 (2000).
8. Ludewig, U., Pusch, M. & Jentsch, T. J. *Nature* **383**, 340–343 (1996).
9. Middleton, R. E., Pheasant, D. J. & Miller, C. *Nature* **383**, 337–340 (1996).
10. Weinreich, F. & Jentsch, T. J. *J. Biol. Chem.* **276**, 2347–2353 (2001).
11. Pusch, M., Ludewig, U., Rehfeldt, A. & Jentsch, T. J. *Nature* **373**, 527–531 (1995).
12. Mindell, J. A., Maduke, M., Miller, C. & Grigorieff, N. *Nature* **409**, 219–223 (2001).
13. Fahlke, C., Rhodes, T. H., Desai, R. R. & George, A. L. *Nature* **394**, 687–690 (1998).
14. Ludewig, U. Thesis, Univ. Göttingen (1996).
15. Maduke, M., Williams, C. & Miller, C. *Biochemistry* **37**, 1315–1321 (1998).
16. Saviane, C., Conti, F. & Pusch, M. *J. Gen. Physiol.* **113**, 457–467 (1999).

Molecular neurobiology

Priming plasticity

Lynn E. Dobrunz and Craig C. Garner

Nerve cells communicate by using chemical messengers, which are released from neurons after a 'priming' step. It seems that priming may be key to controlling the strength of chemical transmission.

The roots of cognition, behaviour, learning and memory are embedded in the brain's intricate network of nerve cells and their specialized points of contact, the synapses. Synapses can convert electrical impulses into chemical signals and back again, as well as modulate the strength of the transmitted signals. This ability to modify the strength of transmission — known as synaptic plasticity — is thought to be the cellular basis of the brain's ability to compute, learn and remember. A goal of many neurobiologists is to understand the molecular basis of synaptic plasticity. On pages 321 and 327 of this issue, two papers from the same groups provide some answers. Schoch *et al.*¹ and Castillo *et al.*² have carried out a series of elegant analyses of a synaptic protein called RIM. In so doing, they show that a key regulatory step in some forms of synaptic plasticity involves 'priming' — preparing the chemical-messenger carriers to release their cargo.

Synapses are made up of the 'bouton' found on the end of the signal-transmitting (presynaptic) neuron; the signal-detecting apparatus on the receiving (postsynaptic) neuron; and the small gap in between (the synaptic cleft; Fig. 1, overleaf). The presynaptic bouton is filled with synaptic vesicles, small membranous vehicles that contain the chemical messengers (neurotransmitters). After a priming step, these vesicles dock at a specialized region of the presynaptic plasma membrane known as the active zone. Electrical impulses arriving at the active zone trigger the fusion of synaptic vesicles with the plasma membrane, releasing their neurotransmitter

into the synaptic cleft (Fig. 1). The neurotransmitter then diffuses across the cleft and activates receptors clustered in the postsynaptic plasma membrane, initiating a new wave of electrical signals in the postsynaptic neuron.

This process of synaptic transmission can be enhanced or weakened by changing the probability that neurotransmitter is released in response to electrical activity in the presynaptic bouton, or by altering the size of the postsynaptic response. These processes together determine the overall strength of synaptic transmission. Synaptic strength is activity dependent — it is modulated by the frequency of electric impulses arriving in the presynaptic bouton and the activity of the postsynaptic neuron. Changes in synaptic strength can be transient, lasting milliseconds to minutes (short-term plasticity³), or long-lasting, persisting for hours, days or months (long-term plasticity⁴).

The molecular underpinnings of synaptic plasticity, in particular long-term plasticity, are thought to involve changes in the composition or activity of presynaptic and/or postsynaptic proteins. Several forms of long-term plasticity involve postsynaptic changes, whereas many forms of short-term plasticity and at least one form of long-term plasticity require presynaptic changes. Presynaptic changes — alterations in the probability of neurotransmitter release — could theoretically occur at several stages in the release process, such as the docking, priming or fusion of synaptic vesicles. Docking involves tethering synaptic vesicles to the plasma membrane at the active zone⁵. Priming prepares docked

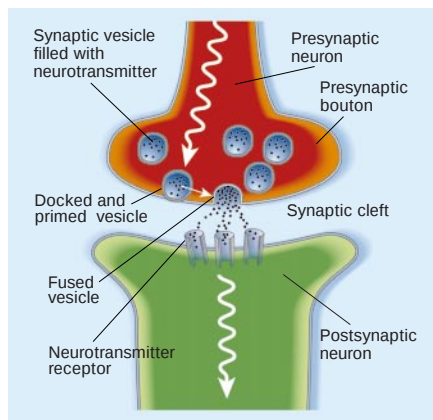


Figure 1 Signalling at a chemical synapse. Chemical synapses are specialized cellular junctions that enable nerve cells to communicate. The presynaptic bouton is filled with vesicles (blue circles) containing neurotransmitter (black dots). These vesicles dock at the plasma membrane, where they undergo a 'priming' step that renders them capable of fusing with the plasma membrane, once an electrical impulse (wavy white line) arrives in the presynaptic bouton. Upon fusion, a vesicle's contents are released into the synaptic cleft. Receptors on the postsynaptic neuron detect the neurotransmitter, and trigger a postsynaptic electrical response. The strength of neurotransmission can be altered, and such 'synaptic plasticity' is thought to underlie the brain's ability to compute, learn and remember. Schoch *et al.*¹ and Castillo *et al.*² now suggest that a synaptic protein called RIM1 α is involved in synaptic plasticity — specifically, through its role in vesicle priming.

vesicles for fusion; they are then held in check until an electrical impulse causes an influx of calcium ions into the neuron, triggering fusion⁵. So changing the activity of certain presynaptic proteins involved in docking, priming or fusion could alter the probability of neurotransmitter release.

Several proteins have emerged from genetic and molecular analyses as possible regulators of presynaptic plasticity. These include synaptotagmin, a calcium sensor involved in vesicle fusion; Munc13, a protein required for priming; and Rab3A, a synaptic-vesicle protein involved in docking. Another synaptic protein that might be involved is RIM, a Rab3A-interacting molecule, whose function has until now been enigmatic. RIM is a modular protein that can bind Munc13 and synaptotagmin as well as Rab3A. These characteristics, as well as the tight and specific association of RIM with the presynaptic active zone, led to the hypothesis that RIM may help to localize its protein partners near their sites of action. Furthermore, its ability to interact with Rab3A indicated that RIM might also be involved in vesicle docking. But genetic studies indicated — at least in nematode worms — that RIM is not essential for the assembly of synapses, nor for the docking or fusion of synaptic vesicles, and that it is

instead involved in synaptic-vesicle priming⁶.

Schoch *et al.*¹ and Castillo *et al.*² have now used a combination of mouse genetics, biochemistry and electrophysiology to investigate the role of RIM in mammals. They analysed synapses from mice in which the gene encoding RIM1 α — one of the two mammalian forms of RIM — was disrupted. The authors find that RIM1 α acts at a vesicle-priming step in mammals, as it does in nematodes. As the probability of vesicle fusion is proportional to the number of primed vesicles, these results suggest that RIM has an essential function in regulating neurotransmitter release.

Excitingly, Schoch *et al.* and Castillo *et al.* also show that RIM1 α is involved in certain forms of short- and long-term synaptic plasticity. In particular, RIM1 α is essential in establishing the long-term potentiation of synaptic strength at certain synapses ('mossy fibre' synapses) in the hippocampal region of the brain⁷. In addition, disrupting RIM1 α causes increases in the long-term depression of synaptic strength at mossy-fibre synapses². It also causes increases in two forms of short-term plasticity at hippocampal CA1 synapses¹. All are forms of presynaptic plasticity. Interestingly, RIM1 α seems to be involved in both Rab3A-dependent and Rab3A-independent contributions to synaptic plasticity^{1,2}, consistent with the idea that RIM, through its many interactions with other proteins, can regulate neurotransmitter release through parallel yet possibly distinct signalling pathways^{1,6,7}.

All in all, it seems that regulation of the priming step of the vesicle cycle has important consequences for synaptic function, by regulating the amount and activity dependence of neurotransmitter release. Thus, vesicle priming might be crucial in determining the fidelity with which presynaptic boutons speak to postsynaptic cells. It remains to be seen whether and how vesicle priming is regulated during normal brain activity, and what function the regulation of priming has in computation, learning and memory. ■

Lynn E. Dobrunz and Craig C. Garner are in the Department of Neurobiology, University of Alabama at Birmingham, Birmingham, Alabama 35294, USA.

e-mails: dobrunz@nrc.uab.edu
garner@nrc.uab.edu

1. Schoch, S. *et al.* *Nature* **415**, 321–326 (2002).
2. Castillo, P. E., Schoch, S., Schmitz, F., Südhof, T. C. & Malenka, R. C. *Nature* **415**, 327–330 (2002).
3. Zucker, R. S. *Curr. Opin. Neurobiol.* **9**, 305–313 (1999).
4. Bliss, T. & Collingridge, G. *Nature* **361**, 31–37 (1993).
5. Südhof, T. C. *Nature* **375**, 645–653 (1995).
6. Koushika, S. P. *et al.* *Nature Neurosci.* **4**, 997–1005 (2001).
7. Lloyd, T. E. & Bellen, H. J. *Nature Neurosci.* **4**, 965–966 (2001).

correction The caption to Corinne Charbonnel's News & Views article of 3 January — "Cosmology: A baryometer is back" (*Nature* **415**, 27–28; 2002) — should have read "primordial abundances of ⁴He (observed in extra-galactic [not galactic] H II regions)".

Daedalus

Drifting continents

The modern world was created by the drifting of the continents, says the theory of plate tectonics. Thus the Atlantic Ocean has widened, and is continuing to do so, as America and Europe drift apart. So, says Daedalus, why have the Atlantic cables not snapped? Presumably because the drift is very slight, a few centimetres a year, and is well within the elastic stretch of a few thousand kilometres of cable.

Nonetheless, a transoceanic cable is a perfect electrical strain gauge. The resistance of the Atlantic cables must be rising as their conductors lengthen and their conductive cross-section falls. Other oceanic cables span the Pacific, the Indian Ocean and so on. They may be shortening as their paths slowly decrease, so their resistance should be falling. Many of these cables were laid in the days of telegraph and low-frequency telephone traffic. It would cost little to take them out of service for measuring, and transfer their burden to modern high-frequency or fibre-optic cables.

So here is a splendid way of checking the theory of continental drift. The world is bound by many cables, and the ships that laid them were navigated carefully and have left detailed logs. The only snag is that the resistance of a single cable cannot be measured. You need at least two, there and back. They must be joined, and their final termination must be coupled by an excellent conductivity bridge.

Sadly, you cannot tell where or how evenly over its length the multiple cable is being stretched. A large extension in a small region (say, where plates are separating) has the same effect as a slight stretch maintained over many kilometres. Daedalus will be pleased by a change in resistance of a few parts per billion. A continuous watch will show whether the movement is slow and steady, or occurs in sudden jerks, possibly with oscillations about each burst of movement.

DREADCO communications engineers are now seeing which cables can be coupled into useful pairs or triangles. To provide a good test of current theories, each such multiple must allow an accurate measure of its round-trip resistance. The engineers may include long pipelines in their list, provided their resistance is also well defined. Railways usually maintain signalling and power currents which would interfere with the measurements. The whole experiment gives a cheap and novel way of testing one of the boldest theories of modern geology. David Jones

Male displays adjusted to female's response

Macho courtship by the satin bowerbird is tempered to avoid frightening the female.

Models of sexual selection generally assume that behavioural courtship displays reflect intrinsic male qualities such as condition, and that males display with maximum intensity to attract females to mate¹. Here we use robotic females in a field experiment to demonstrate that male satin bowerbirds (*Ptilonorhynchus violaceus*) do not always display at maximum intensity — rather, successful males modulate their displays in response to signals from females. Our results indicate that sexual selection may favour those males that can produce intense displays but which know how to modify these according to the female response.

The courtship of the male satin bowerbird consists of a dramatic, coordinated display of feather-puffing, extending the wings suddenly, and running accompanied by a loud buzzing vocalization. The intensity of this display necessitates trade-offs for males and females. Males must display intensely to be attractive, but these displays involve sudden movements and are similar to male–male aggression displays^{2,3} and so may be perceived as threatening by females². Females may be attracted by intense displays that indicate genetic or proximate benefits^{1,4–6}, but they risk being startled repeatedly when threatened by intense displays, which can disrupt or end courtship⁷. This startling may be costly if females cannot efficiently assess potential mates or if they lose the opportunity to mate with more intensely displaying, preferred males. These trade-offs should favour communication because both sexes will benefit if females are not threatened unnecessarily by intense male displays.

During the process of mate selection, the average female satin bowerbird approaches several males for courtship, returns to a

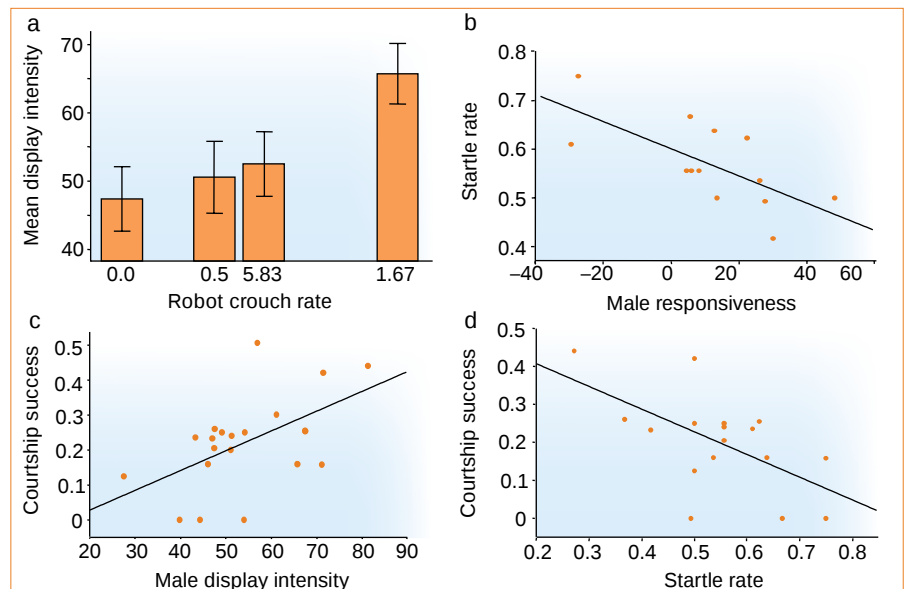


Figure 2 Response of male satin bowerbirds to female signals. **a**, Positive relationship between mean male display intensity ($\pm$ s.e.) and robot crouch rate (ANOVA linear contrast, each male a random block: $F_{1,36}=8.69$, $P<0.006$; display intensity was scored as an index incorporating aspects of male display that affected the threat to females: the degree of feather puffing, the distance run, and male location relative to the female). **b**, Males who are more responsive to female signals are less likely to startle females with intense displays ($n=13$, $r^2=0.47$, $F_{1,11}=9.93$, $P<0.009$; responsiveness is the slope of the male's intensity regressed on robot crouch rate; startle rate was estimated from natural bower activity³ as the average proportion of intense displays that startle females in first courtships between the male and each female he courts). **c**, Males with a higher average display intensity in experimental courtships were more successful in natural courtships ($n=20$, $r^2=0.29$, $F_{1,18}=7.23$, $P<0.015$; male courtship success was measured as the proportion of courted females who mated with the male). **d**, Males who startle females with intense displays less frequently are more successful in natural courtships ($n=17$, $r^2=0.35$, $F_{1,15}=8.10$, $P<0.012$). Further details are available from the authors.

subset of males, and finally copulates with a single male⁸. In each courtship, a female typically 'crouches', lowering herself down and tilting forwards towards the mating position, fluffing her wings when she is fully crouched. Crouching is highly variable, and females may crouch in courtships occurring a week or more before copulation and for males that are not finally chosen as mates. As crouching increases, a female is less likely to be startled by intense male displays, so crouching may

indicate the degree of intensity that she will tolerate without feeling threatened, and signals increasing tolerance during sequential courtships with a male (details of the relationship between crouching and startling are available from the authors). By increasing the intensity of his display as a female's crouching increases, a male should thus be able to maintain his attractiveness without threatening the female.

We tested this hypothesis by measuring the male response to the controlled crouching of robotic female bowerbirds (Fig. 1). Males should respond to increased female crouching by increasing their display intensity, so that high-intensity displays are given when females are likely to be less easily startled. In support of this prediction, we found a highly significant, positive correlation between male display intensity and robot crouch rate (Fig. 2a).

In addition, males that respond more effectively to female signals should threaten females less frequently with intense displays, and we found that there is indeed a negative relation between male responsiveness to robot crouching and the startling of females by male displays in natural courtships (Fig. 2b). When we controlled for responsiveness,



Figure 1 Robotic female satin bowerbird. Robots were capable of three movements, controlled by servo motors: crouching (simultaneous lowering and forward-tilting of the body), looking around during courtship (lateral head movement), and wing-fluffing upon reaching the mating position (lateral wing movement). Robots were enclosed in female satin-bowerbird skins, placed in the bower of the male under test (where females are typically courted⁹) at our field site in Australia, and controlled remotely. The responses of 20 males were tested with robots that crouched from the upright position to the mating position (pictured) at four different rates: fast (3 min total, 1.67 position changes per min), moderate (6 min, 0.83 changes per min), slow (9 min, 0.55 changes per min), and no crouching (0 changes per min), with order randomized between males.

we found a positive relationship between male display intensity and startle rate ($\beta_{\text{response}} = -0.92$, $P < 0.001$; $\beta_{\text{intensity}} = 0.51$, $P < 0.028$; $r^2 = 0.68$, $F_{2,10} = 10.71$, $P < 0.003$).

We predicted that the most successful males would be those who produce high-intensity displays without startling females. Indeed, we found a positive correlation between male display intensity in robot courtships and male success in natural courtships (Fig. 2c), and a negative relationship between startle rate and male courtship success (Fig. 2d), with both factors contributing to male courtship success when considered together ($\beta_{\text{intensity}} = 0.55$, $P < 0.004$; $\beta_{\text{startle}} = -0.57$, $P < 0.003$; $r^2 = 0.65$, $F_{2,14} = 13.23$, $P < 0.0006$).

Our results indicate that although female satin bowerbirds prefer intensely displaying males, successful males do not always display at maximum intensity — instead, they modulate the intensity of their display in response to female signals, to remain attractive without threatening the females. In satin bowerbirds — and perhaps in other species in which variation in sexually selected traits

has not yet been examined in detail — male courtship success may depend on both an attractive display and the intrinsic ability to modify these in response to female signals.

Gail L. Patricelli*, **J. Albert C. Uy†**,
Gregory Walsh‡, **Gerald Borgia***

Departments of *Biology and †Mechanical Engineering, University of Maryland, College Park, Maryland 20742, USA

e-mail: borgia@umail.umd.edu

‡Present address: Department of Ecology, Evolution and Marine Biology, University of California, Santa Barbara, California 93106, USA

1. Andersson, M. *Sexual Selection* (Princeton Univ. Press, Princeton, New Jersey, 1994).
2. Borgia, G. & Presgraves, D. C. *Anim. Behav.* **56**, 1121–1128 (1998).
3. Borgia, G. *Am. Sci.* **83**, 542–548 (1995).
4. Berglund, A., Bisazza, A. & Pilastro, A. *Biol. J. Linn. Soc.* **58**, 385–389 (1996).
5. Borgia, G. in *Sexual Selection and Reproductive Competition in Insects* (eds Blum, M. S. & Blum, N. A.) 27–49 (Academic, New York, 1979).
6. Mateos, C. & Carranza, M. *Behav. Ecol. Sociobiol.* **45**, 235–244 (1999).
7. Borgia, G. *Emu* **95**, 1–12 (1995).
8. Uy, J. A. C., Patricelli, G. L. & Borgia, G. *Proc. R. Soc. Lond. B* **267**, 251–256 (2000).
9. Borgia, G. *Anim. Behav.* **33**, 266–271 (1985).

Laser technology

Measuring huge magnetic fields

Huge magnetic fields are predicted^{1–4} to exist in the high-density region of plasmas produced during intense laser–matter interaction, near the critical-density surface where most laser absorption occurs, but until now these fields have never been measured. By using pulses focused to extreme intensities to investigate laser–plasma interactions⁵, we have been able to record the highest magnetic fields ever produced in a laboratory — over 340 megagauss — by polarimetry measurements of self-generated laser harmonics.

Because harmonics of the laser are generated at the critical-density surface and subsequently propagate isotropically out of the

dense region⁶, we have found that measuring the final polarization of these harmonics is a powerful way to find out the magnitude of the magnetic fields through which they travel. The use of self-generated laser harmonics is particularly convenient because these are produced at precisely the same time as the magnetic fields are generated and propagate so that their **k** vectors are perpendicular to azimuthal magnetic fields in the plasma — which greatly simplifies data interpretation. In our experiments, we use the propagation properties of lower-order harmonics (that is, the third, fourth and fifth harmonics).

These results were obtained with the Vulcan laser system (wavelength 1.054 μm , pulse energy up to 90 J, pulse duration about 1 picosecond). The beam was *p*-polarized and focused to a maximum intensity of $9 \times 10^{19} \text{ W cm}^{-2}$ onto a thin solid target (0.1–1.0 mm). The polarization com-

ponents of the emitted laser harmonics were measured by using high-dynamic-range, charge-coupled-device arrays as detectors.

When an electromagnetic wave propagates in a magnetized plasma with its **k** vector perpendicular to **B**, the extraordinary wave (*x*-wave; that is, with an electric field vector perpendicular to the magnetic field) can experience cut-offs and resonances (Fig. 1a). Cut-offs occur when the plasma index of refraction is equal to zero, and resonances when the index approaches infinity. The *x*-wave is reflected when it encounters a cut-off and is absorbed in a resonance. For example, the cut-offs for the fifth, fourth and third harmonics occur at 460, 340 and 220 megagauss, respectively, for a density of $n_e = 2.4 \times 10^{21} \text{ cm}^{-3}$ (the relativistically corrected critical density). Resonances occur at higher magnetic fields than cut-offs. The ordinary (*o*) wave (with **E** parallel to **B**) is unaffected by the magnetic field — implying that if a field larger than the cut-off field exists in the plasma, then only the ordinary wave is able to propagate to the detector and therefore is the only one observable.

This is what we find for the highest-intensity shots. Figure 1b shows the ratio of *p*-component (*x*-wave) to total emission (*x*-wave plus *o*-wave) for both the third and fourth harmonics for various incident laser intensities. At high intensities, the *x*-wave cut-offs are definitely observed, implying the existence of a minimum magnetic field of 340 megagauss in the plasma; no cut-offs were seen for the fifth harmonic. This indicates that the peak magnetic field is below 460 and above 340 megagauss at intensities of about $9 \times 10^{19} \text{ W cm}^{-2}$. Such fields are more than an order of magnitude larger than any previously observed in the laboratory^{7–9}. These cut-offs were consistently reproducible in our experiments — but only at the highest laser intensities.

The magnitude of the magnetic fields generated in this way could soon approach those needed for testing astrophysical models of neutron stars and white dwarfs¹⁰.

M. Tatarakis*, **I. Watts***, **F. N. Beg***,
E. L. Clark*, **A. E. Dangor***, **A. Gopal***,
M. G. Haines*, **P. A. Norreys†**, **U. Wagnert†**,
M.-S. Wei*, **M. Zepf***, **K. Krushelnick***

*The Blackett Laboratory, Imperial College of Science, Technology and Medicine, London SW7 2BZ, UK

e-mail: m.tatarakis@ic.ac.uk

†Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire OX11 0QX, UK

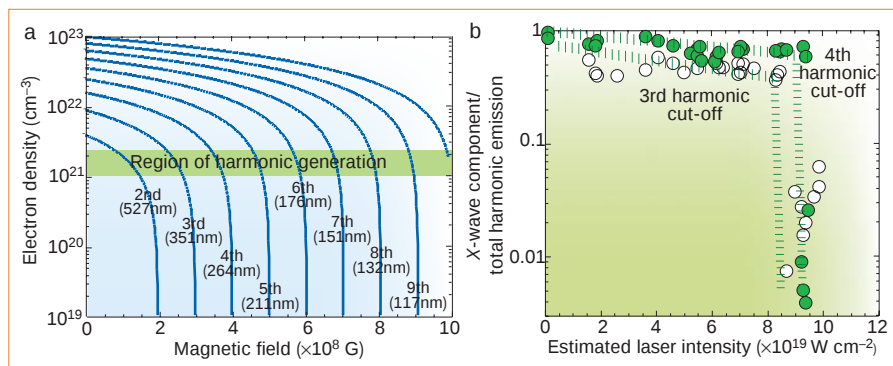


Figure 1 Laboratory measurement of magnetic fields greater than 340 megagauss. **a**, Plot of *x*-wave cut-offs for various harmonics (second, third, and so on) of 1.054- μm radiation in terms of plasma electron density and magnetic field. **b**, *X*-wave/total harmonic emission of third harmonic (hollow circles) and fourth harmonic (filled circles) for a series of laser shots.

1. Wilks, S. C. *et al.* *Phys. Rev. Lett.* **69**, 1383–1386 (1992).
2. Pukhov, A. & Meyer-ter-Vehn, J. *Phys. Rev. Lett.* **76**, 3975–3978 (1996).
3. Mason, R. J. & Tabak, M. *Phys. Rev. Lett.* **80**, 524–527 (1998).
4. Sudan, R. *Phys. Rev. Lett.* **70**, 3075–3078 (1993).
5. Perry, M. D. & Mourou, G. *Science* **264**, 917–924 (1994).
6. Norreys, P. A. *et al.* *Phys. Rev. Lett.* **76**, 1832–1835 (1996).
7. Borghesi, M. *et al.* *Phys. Rev. Lett.* **80**, 5137–5140 (1998).
8. Tatarakis, M. *et al.* *Phys. Rev. Lett.* **81**, 999–1002 (1998).
9. Clark, E. L. *et al.* *Phys. Rev. Lett.* **84**, 670–673 (2000).
10. Lai, D. & Salpeter, E. E. *Astrophys. J.* **491**, 270–285 (1997).

Gate-voltage control of spin interactions between electrons and nuclei in a semiconductor

J. H. Smet*, R. A. Deutschmann†, F. Ertl‡, W. Wegscheider‡§, G. Abstreiter† & K. von Klitzing*

* Max-Planck-Institut für Festkörperforschung, D-70569 Stuttgart, Germany

† Walter Schottky Institut, Technische Universität München, D-85748 Garching, Germany

§ Institut für Experimentelle und Angewandte Physik, Universität Regensburg, D-93040 Regensburg, Germany

Semiconductors are ubiquitous in device electronics, because their charge distributions can be conveniently manipulated with voltages to perform logic operations. Achieving a similar level of control over the spin degrees of freedom, either from electrons or nuclei, could provide intriguing prospects for both information processing and the study of fundamental solid-state physics issues. Here we report procedures that carry out the controlled transfer of spin angular momentum between electrons—confined to two dimensions and subjected to a perpendicular magnetic field—and the nuclei of the host semiconductor, using gate voltages only. We show that the spin transfer rate can be enhanced near a ferromagnetic ground state of the electron system, and that the induced nuclear spin polarization can be subsequently stored and ‘read out’. These techniques can also be combined into a spectroscopic tool to detect the low-energy collective excitations in the electron system that promote the spin transfer. The existence of such excitations is contingent on appropriate electron–electron correlations, and these can be tuned by changing, for example, the electron density via a gate voltage.

Discussions of future information-processing technologies often assign a prominent role to the spin degree of freedom in addition to (or instead of) the charge degree of freedom¹, which is exploited in today’s mainstream electronics. In the short term, this ‘spintronics’ may deliver products with enhanced functionality or improved performance, such as high-speed, high-density non-volatile random access memories: whereas on a much longer timescale, contributions to the very challenging realm of quantum computation^{2,3} have been anticipated. Quantum computation attempts to benefit from correlations and dissipationless transformations of coupled quantum-mechanical systems. The main incentive is a certain degree of parallelism that computational schemes based on such principles bring with them. For example, such schemes offer algorithms for prime factorization⁴ and for exhaustive search⁵; unlike any apparatus based on classical physics, a quantum computer should be able to solve these problems in polynomial time—provided that it can be implemented in a real machine, as energy dissipation is a fundamental source of concern. There has been a wide variety of proposals for practical implementations of rudimentary logic gates in which quantum memory registers—based on any of the abundant two-level systems in physics, like spin-1/2 electrons and nuclei—can be externally manipulated. These proposals range from trap configurations in atom or ion physics⁶, to techniques of nuclear magnetic resonance spectroscopy^{7,8} as used in organic chemistry, to a very bold all-electronic approach for nuclear spin solid-state devices⁹ that would marry the merits of electronics fabrication technology with the virtues of quantum computation¹⁰.

Many of the ideas produced by workers in the spintronics and quantum computing communities may be deemed far out of reach. But they have sparked efforts to develop new ways to accomplish the more fundamental task of controlling and measuring the nuclear spin polarization in solid-state devices, in view of the dearth of existing techniques for locally manipulating nuclear spins. Particularly appealing is the use of mobile objects, like conduction electrons in semiconductors, as mediators to both probe and modify nuclear spins. Gating and optical techniques are able to tailor precisely the population and energy distribution of such electrons, especially

when they are constrained to move in two dimensions—as in quantum wells or field-effect transistors—or even fewer dimensions. The creation of non-equilibrium populations of spin-polarized electrons using coherent polarized light pulses or gating techniques have, for example, already enabled dynamical control of nuclear spins or the electronic generation of net nuclear spin polarization^{11–13}. Progress in this area will rely on experiments specifically geared towards expanding limited knowledge of controlled spin interactions and the microscopic interaction processes that take place between spin systems in such low-dimensional structures. It constitutes the main motivation for the work presented here.

The electrons and nuclear spins in a semiconductor are coupled through the hyperfine interaction. For electrons in the conduction band described by *s*-type Bloch functions the contact hyperfine Fermi interaction prevails¹⁴. It allows the simultaneous reversal of a nuclear and an electronic spin while preserving total spin angular momentum and energy. This flip-flop scattering mechanism may either serve as a relaxation channel to equilibrate the nuclear spin degrees of freedom with the lattice, or as a means of dynamically polarizing and cooling them if the electronic spin system is first brought out of equilibrium. However, in a two-dimensional electron system¹⁵ (2DES) based on GaAs and subjected to a strong perpendicular magnetic field $B_{\perp}$, nuclear spins precess at frequencies some three orders of magnitude below that of the electronic Zeeman gap—which is the energy required to flip a single electronic spin. This large discrepancy normally impedes efficient spin transfer in this regime, so that the nuclear and electronic systems are well isolated. But this barrier for spin transfer can be overcome by taking into account Coulomb correlations inherent in the 2DES: at specific electron densities, which can be controlled by a gate voltage, these correlations contrive low-energy or gapless collective excitations that better match the nuclear spin energy levels, so that efficient channels to convey spin angular momentum between both spin systems can be opened up at these densities.

Here we provide evidence that by varying the carrier density n_{2D} of the 2DES at a fixed field $B_{\perp}$ through a gate voltage, the state of the nuclear spin system can indeed be manipulated, read out and

stored. Spatial variations in the nuclear polarization are generated when the 2DES condenses in a ground state with ferromagnetic order^{16,17} and a rapid thermal equilibration is achieved at carrier densities where gapless spin wave modes¹⁸ (associated with the in-plane magnetic order of finite-size skyrmions, complex spin textural defects^{19–21}) are believed to exist. Depletion of the 2DES interrupts flip-flop scattering, and the nuclear spin configuration is stored. Dwelling at other electron densities, where the energy conservation rule for flip-flop scattering can not be fulfilled, also quenches the hyperfine interaction and the nuclear spin orientation is maintained. The degree of nuclear spin polarization acts back on the electronic system and relocates the ferromagnetic ordered ground

state to higher fields, so that its position reveals the degree of nuclear spin polarization.

The Ising ferromagnetic state

The quantization of the electron cyclotron motion in a field $B_{\perp}$ and the Zeeman coupling of $B_{\perp}$ to the electron spin degree of freedom discretizes the energy spectrum of a 2DES into a ladder of spin-split Landau levels. Each level is classified by its orbital radius and spin quantum number, and possesses a macroscopic degeneracy per unit area $d = B_{\perp}/\Phi_0$ where Φ_0 is the flux quantum¹⁵. The filling factor $\nu = n_{2D}/d$ denotes the number of filled Landau levels. A similar effect occurs by virtue of the Coulomb interaction when all electrons reside in the lowest Landau level. When it is half filled, composite fermions²²—quasiparticles assembled from one electron and two flux quanta—make up a metallic state and a fan of composite-fermion Landau levels develops away from filling factor $\nu = 1/2$. If an integer number p of electron Landau levels is filled, the 2DES condenses in an integer quantum Hall state²³ with its vanishing conductance along the current direction and quantized Hall conductance. Analogously, if an integer number q of composite-fermion Landau levels is occupied (equivalent to fractional fillings of the lowest electron Landau level of the form $\nu = q/(2q \pm 1)$), the fractional quantum Hall effect^{24,25} ensues.

When two electron or composite-fermion Landau levels simultaneously approach the chemical potential, correlations frequently force the 2DES to order in a ferromagnetic ground state^{16,17,26–29} similar to those of conventional low-dimensional ferromagnets. If the Landau levels have opposite spin quantum numbers and different cyclotron energies, the exchange energy cost for spin misalignment forces those particles that are to be distributed among the nearly coincident Landau levels to take on identical spin orientation, either along or opposite to the Zeeman field. It induces an Ising-like ferromagnetic phase transition with easy-axis anisotropy between quantum Hall states with distinct spin configurations as one Landau level overtakes the other. It is heralded in the transport properties of the 2DES by the disappearance of the quantum Hall effect. This first-order phase transition is accompanied by hysteresis and extra dissipation due to the domain morphology, which acts as an additional source of backscattering for current-carrying quasiparticles. An example is shown in Fig. 1 for the spin-unpolarized and fully spin-polarized $\nu = 2/3$ fractional quantum Hall states when two composite-fermion Landau levels are completely occupied, and the spin-down state of the lowest level and the spin-up state of the second level are brought close to degeneracy (this and all other experiments were performed at the base temperature of a 20 mK dilution refrigerator). The quantized kinetic energy of composite fermions originates from the Coulomb interaction energy E_C , and grows with $\sqrt{n_{2D}}$ or equivalently $\sqrt{B_{\perp}}$ at fixed filling, whereas the Zeeman splitting E_Z goes linearly. This phase transition can thus be conveniently driven by tuning the carrier density and occurs at a critical ratio of both energy scales.

In the presence of spin orientation, the hyperfine interaction also causes nuclear (B_N) and electronic (B_e) magnetic fields that act respectively on the electronic and nuclear spins, as manifested for example by Overhauser and Knight shifts in their energy spectrum¹⁴. The hyperfine effective nuclear field B_N only alters the electronic Zeeman energy, so that $E_Z/E_C \propto (B_{\perp} + B_N)/B_{\perp}^{1/2}$. From the external field $B_{\perp}$ or density for which E_Z/E_C exceeds the critical value that triggers the Ising ferromagnetic phase transition, the degree of nuclear spin orientation B_N can be extracted. A gradual polarization of the nuclear spin system as time progresses should leave a ‘fingerprint’ in the source–drain resistance R_{SD} as the phase transition is relocated, provided that the fixed working point (n_{2D} , $\nu \approx 2/3$) at which R_{SD} is monitored is chosen skilfully.

Flipping of nuclear spins

At the transition, the sample is unable to reach instantaneously its

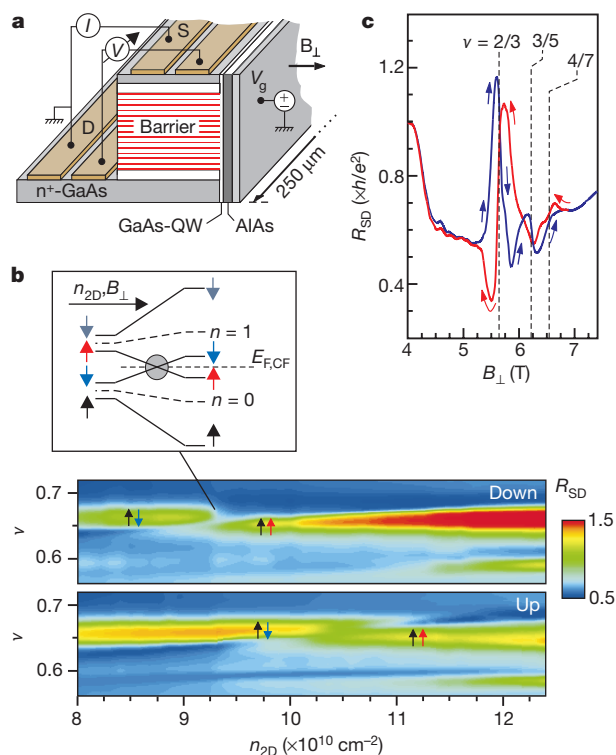


Figure 1 Ferromagnetic phase transition with easy-axis anisotropy between the spin-unpolarized and fully spin-polarized $2/3$ fractional quantum Hall states. **a**, The device is a field-effect transistor on the cleaved edge of a GaAs (001) substrate with a channel length of $3 \mu\text{m}$ and width of approximately $250 \mu\text{m}$. In a first growth sequence along the (001) direction, a superlattice barrier is sandwiched between two conducting n^+ -GaAs layers that form the source (S) and drain (D) contacts. Subsequently, the sample is cleaved *in situ* and a thin GaAs layer (GaAs-QW) is deposited on the freshly exposed (110) surface. It will host the 2DES. It is capped by an AlAs barrier and an n^+ -GaAs layer that serves as the gate. A positive gate voltage V_g with respect to source and drain induces the 2DES electrostatically at the GaAs/AlAs heterointerface. For this geometry, a plateau appears in the source–drain resistance R_{SD} when the 2DES condenses in a quantum Hall state (or a maximum if it is not fully developed)¹⁷. In between two such states, the non-zero source–drain conductivity partly short-circuits the Hall voltage and therefore R_{SD} drops and displays a minimum. **b**, Two-dimensional graphs of R_{SD} (colour) versus carrier density n_{2D} and filling factor ν near $\nu = 2/3$ in units of h/e^2 . The data acquisition is performed during downward (top) and upward (bottom) sweeps of the magnetic field. The hysteretic phase transition is signalled by the disappearance of the $\nu = 2/3$ quantum Hall maximum. Inset: at the coincidence of the composite-fermion spin-up Landau level with orbital index $n = 1$ and the composite-fermion spin-down Landau level with index $n = 0$, the unpolarized $2/3$ fractional quantum Hall state (black and blue arrows indicate the spin orientation of the participating filled levels) vanishes and the fully spin-polarized $2/3$ fractional quantum Hall state (black and red arrows) emerges. ($E_{F,CF}$ is the Fermi level.) **c**, Example of hysteresis in an R_{SD} trace when reversing the magnetic field sweep direction.

stationary or thermodynamic equilibrium state. Figure 2 illustrates the delayed response of the resistance after a magnetic field sweep or change in gate voltage took place. We believe that the sample steadily approaches equilibrium with the aid of the hyperfine interaction, and that the initial time dependence of the resistance mainly reflects a gradual change in the nuclear spin polarization. The non-equilibrium character may originate from either the electronic system itself or the nuclear lattice. In the ferromagnetic state, energy barriers control the dynamics of domain walls, and the electronic system may be trapped in a metastable state due to a local free energy minimum. The growth of domains requires spin flips that can be mediated by the underlying nuclear lattice. Because the degree of electron spin polarization can be drastically modified with filling factor and thus magnetic field or gate voltage, the nuclear spin system may have been brought out of equilibrium by sweeping the external field or carrier density. The involvement of nuclear spins has been previously inferred from resistively detected NMR in this regime^{17,30}. Other evidence is depicted in Fig. 2. During removal of all charge carriers from the 2DES for a short interval, the resistance change halts³¹. Subsequently it continues as if depletion had not taken place. Only the nuclear spins can be invoked as the storage medium for the domain configuration in the absence of charge carriers. We will now substantiate the above claims—but we will first illustrate how the nuclear spin system may be returned to a well-defined, reproducible initial state ('reset') irrespective of the previous history of the sample, an essential ability for a systematic investigation of these phenomena.

'Reset' procedure for the nuclei

At Landau level filling factor $\nu = 1$, quantum Hall systems exhibit Heisenberg-like isotropic itinerant ferromagnetic order in their ground state²⁰. The Coulomb exchange energy penalty for reversed spins exactly complete spin alignment of the free electrons, even in the limit of vanishing Zeeman coupling strength. Furthermore, at sufficiently weak Zeeman coupling it turns finite-size skyrmions or anti-skyrmions^{20,32} (instead of single spin flip excitations) into the lowest-energy charged excitations of this incompressible ground state. These circularly symmetric spin textural defects have their spins aligned with the Zeeman field at the perimeter, gradually reversed towards their centre to keep the exchange cost to a minimum, and possess non-zero in-plane spin components with a vortex-like configuration at intermediate distances. They accommodate precisely one extra unit of charge, and yet flip numerous spins depending on their spatial extent. This spatial extent is set by

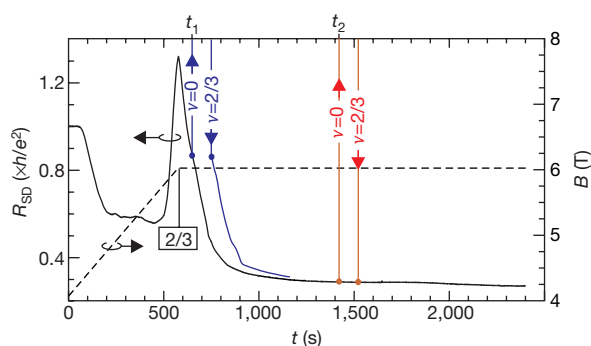


Figure 2 Time dependence of R_{SD} near filling factor $\nu = 2/3$ after a field sweep. The magnetic field sweep starts from $\nu = 1$ and is interrupted at $\nu = 2/3$. R_{SD} continues to change. The experiment is then repeated, but at times t_1 (blue curve) and t_2 (red curve) all charge carriers are removed for 90 s by reducing the gate voltage to zero. When the original carrier density is restored, the resistance continues its descent from nearly the same value as if the time interval during which the sample was fully depleted never took place.

the competition between the Zeeman coupling, which attempts to reduce the number of flipped spins by covering a small area, and the Coulomb interaction, which favours a uniform charge distribution across a large area. At exactly $\nu = 1$, skyrmions freeze out and disappear at low temperatures. However, when moving away from filling factor $\nu = 1$, for each unit of charge added or subtracted from the 2DES a skyrmion or anti-skyrmion is generated. At a finite density of skyrmions, it has been proposed that they crystallize in the sufficiently dilute limit³³. The classical energy of an isolated skyrmion is invariant for translations in real space and uniform rotations in spin space about the axis defined by the Zeeman field. In a crystalline configuration the spin rotational and translational symmetries are broken, and these degrees of freedom give rise to gapless low-energy collective excitations in the theory¹⁸. Even if the crystal melts at higher skyrmion densities, these modes presumably remain present in an overdamped form. The gapless spin wave Goldstone mode then provides an efficient channel for flip-flop scattering, as the energy conservation criterion for the Fermi contact interaction can be met. This mode has been invoked as one possible explanation to account for both the rapid nuclear spin relaxation rate in optically pumped NMR experiments^{21,34} and the enormous enhancement of the specific heat³⁵ due to the entropy of the nuclei near (but away from) exact filling $\nu = 1$.

Here we exploit these previously reported short nuclear spin relaxation times T_1 in the skyrmion regime to bring the nuclear spin

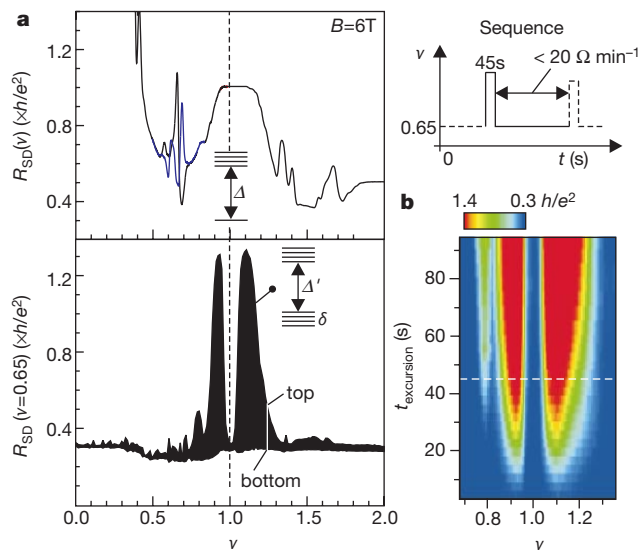


Figure 3 Time dependence of R_{SD} after a short excursion from filling factor $\nu = 0.65$ to filling factor ν at fixed magnetic field. **a**, Top panel: R_{SD} as a function of ν obtained by sweeping the carrier density down (blue curve) or up (black curve) at fixed magnetic field. Bottom panel: the measurement sequence executed to obtain the data in this panel is depicted in the inset on the right. R_{SD} is left to relax at $\nu = 0.65$ until $dR_{SD}/dt < 20 \Omega \text{ min}^{-1}$ (dashed line). Then, a new measurement sequence starts (solid line). First, a $t_{\text{excursion}} = 45 \text{ s}$ excursion to filling factor ν is performed. The resistance upon return to $\nu = 0.65$ is recorded (envelope top) and its subsequent time dependence is plotted along the vertical axis until $dR_{SD}/dt < 20 \Omega \text{ min}^{-1}$ (envelope bottom). The same procedure is then repeated for a different filling factor (dashed line on the right of the insert describing the measurement sequence). The data are also featureless from $\nu = 2$ up to the maximum reachable filling factor of 3.3 (limited by the gate leakage current, data not shown). The energy diagrams indicate schematically that only near (but away from) $\nu = 1$ low-energy collective excitation modes (δ) exist, compatible with the energy spectrum of the nuclei. Δ' and Δ are the energy required to create a single finite-size skyrmion. At exact filling factor $\nu = 1$, the large gap Δ prevents flip-flop scattering. **b**, Same as before, but with $t_{\text{excursion}}$ as an additional parameter. R_{SD} at $t = 0 \text{ s}$ upon return is plotted (red corresponds to high values, blue to low values). Dashed white line corresponds to the envelope top in **a** for $t_{\text{excursion}} = 45 \text{ s}$.

system into thermal equilibrium with the GaAs lattice for a specific filling factor near $\nu = 1$. The history-dependent nuclear spin polarization acquired near the $\nu = 2/3$ ferromagnetic phase transition is erased. Figure 3a illustrates the recovery of the high values of source–drain resistance R_{SD} near the $\nu = 2/3$ ground-state spin transition after a short excursion ($t_{\text{excursion}} = 45$ s) of the sample at fixed magnetic field to filling factors ν close to, but distinct from, $\nu = 1$. The difference in R_{SD} at $\nu = 0.65$, ΔR_{SD} , between its value at $t = 0$ directly upon return to $\nu = 0.65$ from filling factor ν and its value at large t (defined as $dR_{SD}/dt < 20 \Omega \text{ min}^{-1}$) is a measure of the nuclear spin–lattice relaxation time T_1 , and provides indirect information on the available electronic many-particle energy states at filling factor ν . The data exhibit maxima near $\nu = 0.92$ and 1.1 , and reproduce qualitatively the behaviour of T_1 obtained from optically pumped NMR experiments^{21,34}. At $\nu = 1$, in the absence of skyrmions and the associated low-energy collective modes, the nuclear spin polarization state is left unaltered and stored. By including $t_{\text{excursion}}$ as an additional parameter in Fig. 3b, values of T_1 of around 20 s at $\nu = 0.9$ are estimated from the saturation of $\Delta R_{SD}(t_{\text{excursion}})$, in

agreement with ref. 34. Thus, staying for more than 40 s at $\nu = 0.9$ apparently suffices to reset the nuclear spin lattice to a well-defined state and wipe out any previous history.

Putting it all together

This reset procedure is the key to the corroboration of our initial assertion that the time dependence of R_{SD} and the hysteretic behaviour at the ferromagnetic phase transition (Figs 1 and 2) mainly originate from a gradual polarization change of the nuclear

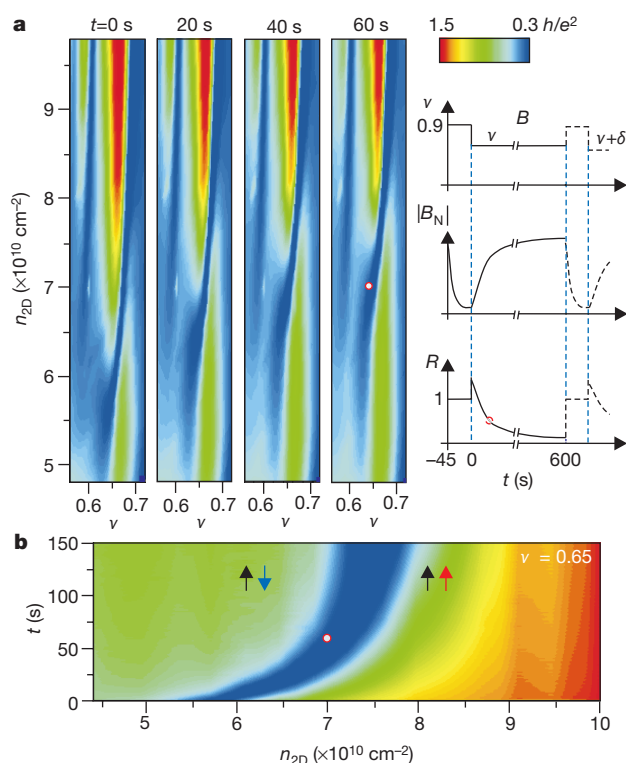


Figure 4 Time progression of the ferromagnetic phase transition. The time dependence of R_{SD} is acquired for closely spaced coordinates in the (ν, n_{2D}) -plane by following the measurement sequence depicted on the right. Before recording the time dependence at a certain (ν, n_{2D}) -pair, the filling factor is set to $\nu = 0.9$ for 45 s. The nuclear spin system of the host GaAs crystal is allowed to relax to its thermal equilibrium state for this filling factor due to the enhanced flip-flop scattering rate under these conditions. It ensures a well-defined, history-independent starting point. After recording R_{SD} for 600 s, either the magnetic field or gate voltage is swept to the next (ν, n_{2D}) -pair and the above procedure is repeated (dashed line in the inset plotting the measurement sequence). Several sections through the three-dimensional parameter space (t, ν, n_{2D}) are displayed. The resistance values are colour coded (red, high resistance values; blue, small resistance values). **a**, From left to right: $R_{SD}(\nu, n_{2D})$ at time $t = 0, 20, 40, 60$ s. **b**, $R_{SD}(n_{2D}, t)$ at fixed filling factor $\nu = 0.65$. This graph and similar cuts through the data at nearby filling factors and extended to longer times t are an aid to selecting a suitable working point at which the source–drain resistance is monitored for the experiments plotted in Figs 3, 5–7. Working points meet the following criteria: monotonic behaviour and maximum swing of R_{SD} for the variation in B_N relevant to the experiments. The white dot with red border is the data point for $(t = 60 \text{ s}, \nu = 0.65, n_{2D} = 7 \times 10^{10} \text{ cm}^{-2})$.

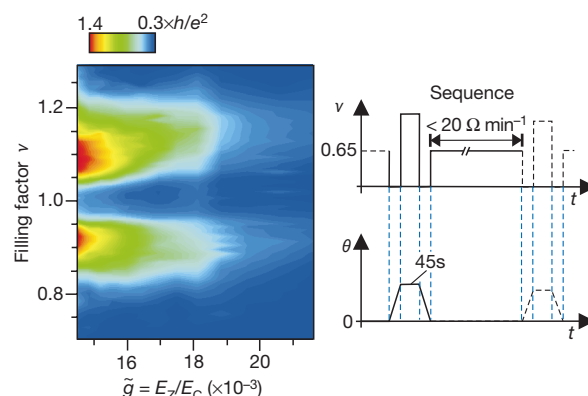


Figure 5 The ratio E_Z/E_C can be changed by tilting the sample *in situ* with respect to the axis of the superconducting magnet. (E_C is the Coulomb interaction energy, and E_Z is the Zeeman splitting.) This is because the Coulomb interaction scales with the perpendicular component of the magnetic field only, whereas the Zeeman energy is determined by the total field. The magnet produces a fixed field of 6 T during the entire experiment. The nuclear spin relaxation rate near $\nu = 1$ is investigated as in Fig. 3, but as a function of E_Z/E_C by executing the following measurement sequence: R_{SD} is left to relax at $\nu = 0.65$ until $dR_{SD}/dt < 20 \Omega \text{ min}^{-1}$. Subsequently, all charge carriers are removed and the sample is rotated with a nearly frictionless mechanism from $\theta = 0$ to $\theta = \alpha$ within a time of 40 s, independent of the final angle set. θ is defined as the angle between the magnetic field and the direction perpendicular to the 2DES. In the absence of charge carriers during rotation, the transfer of spin angular momentum is interrupted and the nuclear spin configuration is retained as proven in Fig. 2. The system is then left to relax for 45 s at a filling factor ν near 1. The 2DES is depleted once more for 40 s. During this time, the sample is reoriented perpendicular to the field. The original filling factor $\nu = 0.65$ is restored, and R_{SD} is acquired immediately afterwards. R_{SD} is allowed to relax, so that a new sequence at different ν or angle can be started. The procedure ensures that E_Z/E_C is always the same at $\nu = 0.65$, so that only the influence of a variation in E_Z/E_C on the skyrmion physics near $\nu = 1$ is tested. The skyrmion size gradually shrinks, and eventually the square lattice configuration becomes unstable and the spin wave Goldstone modes vanish according to theory¹⁸. The hyperfine interaction is quenched, and R_{SD} experiences essentially no change after the 45 s excursion to a filling factor near $\nu = 1$.

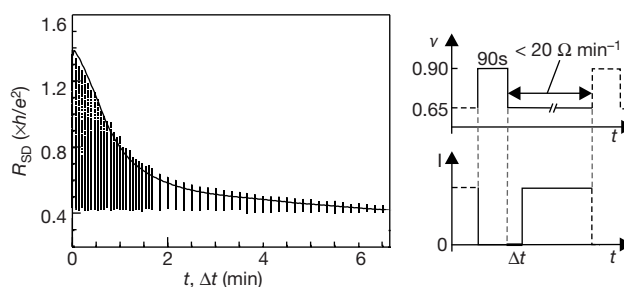


Figure 6 Time behaviour of R_{SD} in the absence of current flow. Thick solid line; $R_{SD}(t)$ at $\nu = 0.65$ and $B = 5.8 \text{ T}$ after equilibrating the nuclear spin system for 90 s at $\nu = 0.9$. Vertical lines; after an excursion of 90 s to $\nu = 0.9$ to return to a well-defined state of the nuclear spin system, the current remains switched off for a time Δt . Subsequently, the resistance is recorded along the vertical axis at abscissa Δt until dR_{SD}/dt is smaller than $20 \Omega \text{ min}^{-1}$. The same procedure is then repeated for different values of Δt .

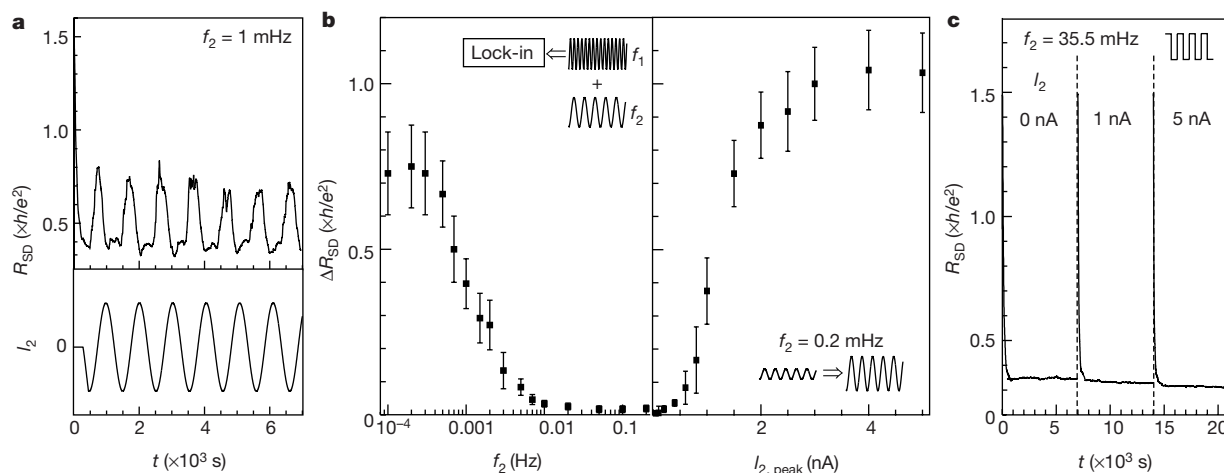


Figure 7 Current-induced displacement of the ferromagnetic phase transition. The following operations are performed to study the influence of current flow. After equilibration of the nuclear spins at $\nu = 0.9$ for 90 s, R_{SD} is monitored with time at $\nu = 0.65$ in the presence of an alternating current with frequency $f_1 = 5.36$ Hz and amplitude $I_1 = 1$ nA. An alternating current with variable amplitude I_2 and frequency f_2 , outside the detection band of the lock-in amplifier utilized to record R_{SD} , is added after 500 s. During the first 500 s the initial rapid drop in R_{SD} took place, associated with the relocation of the phase transition to higher fields due to the change in the net nuclear spin polarization. R_{SD} is then measured for at least 6 periods of this extra alternating current component I_2 . $B_{\perp}$ is left fixed at all times (5.8 T). **a**, R_{SD} obtained in this fashion for $f_2 = 1$ mHz and $I_2 = 1.5$ nA.

spin lattice. R_{SD} is recorded in the three-dimensional parameter space (ν, n_{2D}, t) for $\nu = 0.56 \dots 0.72$, $n_{2D} = 4.8 \times 10^{10} \dots 9.8 \times 10^{10} \text{ cm}^{-2}$ and $t = 0 \dots 600$ s by repeatedly equilibrating the nuclear spin lattice at $\nu = 0.9$ for any pair (ν, n_{2D}). Figure 4 depicts two-dimensional cross-sections through the data, collected in this fashion, either at a fixed time $t = 0, 20, 40$ and 60 s (Fig. 4a) or a fixed filling factor $\nu = 0.65$ (Fig. 4b). The ferromagnetic phase transition progressively shifts to larger carrier densities, or equivalently magnetic fields, as time passes. A time-dependent $B_N(t)$ then provides a natural interpretation of the data, since the critical ratio $E_Z/E_C \propto [B_{\perp} + B_N(t)]/B_{\perp}^{1/2}$, at which the ferromagnetic phase transition due to the coincidence of two composite-fermion Landau levels occurs, is attained at a time-dependent external critical field. The relocation to higher $B_{\perp}$ implies a depolarization of the electronic system and an upward polarization of the nuclear spins, inducing a ΔB_N opposite to the external $B_{\perp}$. From the data at time $t = 0$ (Fig. 4a) and $t \gg 0$ (for example Fig. 1b, no equilibration of the nuclear spin lattice at $\nu = 0.9$), we extract a change in the effective nuclear field of up to $\Delta B_N \approx -0.8$ T on a timescale of a few minutes only. The hysteresis observed in Fig. 1b and c may thus in part be accounted for as originating from a different degree of nuclear spin orientation accumulated in the course of the complex path followed by field and gate voltage to map the resistance in the (n_{2D}, ν)-plane.

This apparently efficient transfer of spin angular momentum to the nuclear lattice near $\nu = 2/3$ strongly suggests the availability of low-energy collective electronic excitations specific to the ferromagnetic state. At present, we can only speculate on their microscopic nature, as ferromagnetism in the fractional quantum Hall regime has not so far been treated theoretically. These collective excitations are probably generated at and localized to the domain boundaries only, where they are presumably responsible for gradually reversing the spin in a Bloch wall between adjacent domains of distinct spin configurations. The procedures developed here may more generally be applied as a spectroscopy tool to unveil the absence or presence of gapless or low-energy collective modes. A

R_{SD} partially returns to its high value immediately after equilibration of the nuclear spin system at $\nu = 0.9$. This is reversible. Its low value is recovered when the sign of the current changes. This experiment suggests that current-driven dynamic nuclear spin polarization erases part of the net nuclear spin polarization when current flows in a particular direction. In the opposite direction the nuclear spin polarization is enhanced. **b**, Left panel: amplitude of the periodic change ΔR_{SD} as a function of f_2 at fixed $I_2 = 1.5$ nA. Error bars indicate the spread in the observed change. Right panel: ΔR_{SD} as a function of I_2 for a fixed frequency $f_2 = 0.2$ mHz. **c**, To verify that the additional dissipation due to I_2 can be excluded as the origin of ΔR_{SD} , $R_{SD}(t)$ is measured for a square wave with $f_2 = 35.5$ Hz and I_2 up to 5 nA.

sophisticated example is depicted in Fig. 5. By selectively enhancing the Zeeman energy over the Coulomb energy scale through the addition of an in-plane magnetic field near $\nu = 1$, the skyrmions shrink in size and approach the single spin flip excitation limit. According to theory¹⁸, the square lattice configuration should become unstable and the spin wave Goldstone modes should vanish—apparently in agreement with this experiment.

Irrelevance of external current flow

It is a common premise that dynamic nuclear spin polarization of this magnitude demands current flow, in view of the large discrepancy between the electron number and the number of nuclei involved. Contacts serve as source or sink of electronic spin in this open system. Yet current flow is not compulsory for the progressive relocation to higher carrier densities of the ground-state spin transition in Fig. 4. Figure 6 illustrates that, even in the absence of any source–drain current, R_{SD} evolves in time along nearly the same curve. A non-equilibrium in the electronic system—due to the domain morphology in the ferromagnetic state—and/or the nuclear spin system—as a result of different degrees of electron spin polarization and corresponding Knight shifts between $\nu = 0.65$ and $\nu = 0.9$ —is the suggested impetus for exchange of spin angular momentum; the direction of this exchange is controlled by the filling factor. An extra alternating current component I_2 can force the phase transition to backtrack to smaller carrier densities, but, as shown in Fig. 7, this current-induced dynamic nuclear polarization occurs only if the frequency f_2 is below 0.01 Hz, whereas the alternating current I_1 used to probe R_{SD} has a frequency f_1 of 5.36 Hz. This low cut-off frequency is consistent with the previously estimated characteristic response time (a few minutes) of the hyperfine interaction near the ferromagnetic state. □

Received 14 September; accepted 22 November 2001.

1. Prinz, G. A. Magneto-electronics. *Science* **282**, 1660–1663 (1998).
2. Bennett, C. H. & DiVincenzo, D. P. Quantum information and computation. *Nature* **404**, 247–255 (2000).
3. Steane, A. Quantum computing. *Rep. Prog. Phys.* **61**, 117–173 (1998).

4. Ekert, A. & Jozsa, R. Quantum computation and Shor's factoring algorithm. *Rev. Mod. Phys.* **68**, 733–753 (1996).
5. Grover, L. K. Quantum mechanics helps in searching for a needle in a haystack. *Phys. Rev. Lett.* **79**, 325–328 (1997).
6. Cirac, J. I. & Zoller, P. Quantum computations with cold trapped ions. *Phys. Rev. Lett.* **74**, 4091–4094 (1995).
7. Cory, D. G., Fahmy, A. F. & Havel, T. F. Ensemble quantum computing by nuclear magnetic resonance spectroscopy. *Proc. Natl Acad. Sci. USA* **94**, 1634–1639 (1997).
8. Gershenfeld, N. A. & Chuang, I. L. Bulk spin resonance quantum computation. *Science* **275**, 350–356 (1997).
9. Kane, B. E. A silicon based nuclear spin quantum computer. *Nature* **393**, 133–137 (1998).
10. Loss, D. & DiVincenzo, D. P. Quantum computation with quantum dots. *Phys. Rev. A* **57**, 120–126 (1998).
11. Kikkawa, J. M. & Awschalom, D. D. All optical magnetic resonance in semiconductors. *Science* **287**, 473–476 (2000).
12. Kawakami, R. K. *et al.* Ferromagnetic imprinting of nuclear spins in semiconductors. *Science* **294**, 131–134 (2001).
13. Wald, K. W., Kouwenhoven, L. P., McEuen, P. L., van der Vaart, N. C. & Foxon, C. T. Local dynamic nuclear polarisation using quantum point contacts. *Phys. Rev. Lett.* **73**, 1011–1014 (1994).
14. Abragam, A. *The Principles of Nuclear Magnetism* 354–423 (Oxford Univ. Press, Oxford, 1961).
15. Ando, T., Fowler, A. B. & Stern, F. Electronic properties of two-dimensional systems. *Rev. Mod. Phys.* **54**, 437–672 (1982).
16. Eom, J. *et al.* Quantum Hall ferromagnetism in a two-dimensional electron system. *Science* **289**, 2320–2323 (2000).
17. Smet, J. H., Deutschmann, R. A., Wegscheider, W., Abstreiter, G. & von Klitzing, K. Ising ferromagnetism and domain morphology in the fractional quantum Hall regime. *Phys. Rev. Lett.* **86**, 2412–2415 (2001).
18. Côté, R. *et al.* Collective excitations, NMR, and phase transitions in skyrme crystals. *Phys. Rev. Lett.* **78**, 4825–4828 (1997).
19. Das Sarma, S. & Pinczuk, A. (eds) *Perspectives on Quantum Hall Effects* (Wiley, New York, 1996).
20. Girvin, S. M. in *Topological Aspects of Low Dimensional Systems* (eds Comtet, A., Jolicœur, T., Ouvry, S. & David, F.) 53–175 (Springer, Berlin, and EDP Sciences, Les Ulis, 1999).
21. Barrett, S. E., Dabbagh, G., Pfeiffer, L. N., West, K. W. & Tycko, R. Optically pumped NMR evidence for finite-size skyrmions in GaAs quantum wells near Landau level filling $\nu = 1$. *Phys. Rev. Lett.* **74**, 5112–5115 (1995).
22. Heinonen, O. (ed.) *Composite Fermions* (World Scientific, Singapore, 1998).
23. von Klitzing, K., Dorda, G. & Pepper, M. New method for high-accuracy determination of the fine-structure constant based on the quantized Hall resistance. *Phys. Rev. Lett.* **45**, 494–497 (1980).
24. Tsui, D. C., Stormer, H. L. & Gossard, A. C. Two-dimensional magnetotransport in the extreme quantum limit. *Phys. Rev. Lett.* **48**, 1559–1562 (1982).
25. Laughlin, R. B. Anomalous quantum Hall effect: an incompressible quantum fluid with fractional charge excitations. *Phys. Rev. Lett.* **50**, 1395–1398 (1983).
26. Jungwirth, T. & MacDonald, A. H. Pseudospin anisotropy classification of quantum Hall ferromagnets. *Phys. Rev. B* **63**, 035305–035314 (2001).
27. Cho, H. *et al.* Hysteresis and spin transitions in the fractional quantum Hall effect. *Phys. Rev. Lett.* **81**, 2522–2525 (1998).
28. Piazza, V. *et al.* First order phase transition in a quantum Hall ferromagnet. *Nature* **402**, 638–641 (1999).
29. De Poortere, E. P., Tutuc, E., Papadakis, S. J. & Shayegan, M. Resistance spikes at transitions between quantum Hall ferromagnets. *Science* **290**, 1546–1549 (2000).
30. Kronmüller, S. *et al.* New type of electron nuclear-spin interaction from resistively detected NMR in the fractional quantum Hall effect regime. *Phys. Rev. Lett.* **82**, 4070–4073 (1999).
31. Hashimoto, K., Muraki, K., Saku, T. & Hirayama, Y. Longitudinal resistance anomaly around the 2/3 filling factor observed in a GaAs/AlGaAs single heterostructure. *Physica B* **298**, 191–194 (2001).
32. Sondhi, S. L., Karlhede, A., Kivelson, S. A. & Rezayi, E. H. Skyrmions and the crossover from the integer to the fractional quantum Hall effect at small Zeeman energies. *Phys. Rev. B* **47**, 16419–16426 (1993).
33. Brey, L., Fertig, H. A., Côté, R. & MacDonald, A. H. Skyrme crystal in a two-dimensional electron gas. *Phys. Rev. Lett.* **75**, 2562–2565 (1995).
34. Tycko, R., Barrett, S. E., Dabbagh, G., Pfeiffer, L. N. & West, K. W. Electronic states in gallium arsenide quantum wells probed by optically pumped NMR. *Science* **268**, 1460–1463 (1995).
35. Bayot, V., Grivei, E., Melinte, S., Santos, M. B. & Shayegan, M. Giant low temperature heat capacity of GaAs quantum wells near Landau level filling $\nu = 1$. *Phys. Rev. Lett.* **76**, 4584–4587 (1996).

Acknowledgements

We thank A. Stern and L. Brey for discussions, and M. Bichler for technological assistance. This work was supported by the German Ministry of Science and Education (BMBF) and the German Science Foundation (DFG).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.H.S. (e-mail: j.smet@fkf.mpg.de).

X-ray structure of a ClC chloride channel at 3.0 Å reveals the molecular basis of anion selectivity

Raimund Dutzler*, Ernest B. Campbell*, Martine Cadene†, Brian T. Chait† & Roderick MacKinnon*

Howard Hughes Medical Institute, * Laboratory of Molecular Neurobiology and Biophysics, and † Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, Rockefeller University, 1230 York Avenue, New York, New York 10021, USA

The ClC chloride channels catalyse the selective flow of Cl[−] ions across cell membranes, thereby regulating electrical excitation in skeletal muscle and the flow of salt and water across epithelial barriers. Genetic defects in ClC Cl[−] channels underlie several familial muscle and kidney diseases. Here we present the X-ray structures of two prokaryotic ClC Cl[−] channels from *Salmonella enterica* serovar *typhimurium* and *Escherichia coli* at 3.0 and 3.5 Å, respectively. Both structures reveal two identical pores, each pore being formed by a separate subunit contained within a homodimeric membrane protein. Individual subunits are composed of two roughly repeated halves that span the membrane with opposite orientations. This antiparallel architecture defines a selectivity filter in which a Cl[−] ion is stabilized by electrostatic interactions with α -helix dipoles and by chemical coordination with nitrogen atoms and hydroxyl groups. These findings provide a structural basis for further understanding the function of ClC Cl[−] channels, and establish the physical and chemical basis of their anion selectivity.

Potassium, sodium, calcium and chloride ions are used ingeniously by living systems in the performance of fundamental cellular tasks. Through the action of ion pumps, a large fraction of a cell's metabolic energy is spent establishing transmembrane ion gradients. These gradients, through the action of ion channels, are used to produce electrical signals, activate signal transduction pathways, regulate cell volume, and mediate fluid and electrolyte transport. To carry out these tasks, an ion channel has to be selective, that is, permit only certain ionic species to flow through its pore. Much is known about a very large family of cation channels, which includes the K⁺, Na⁺, Ca²⁺, cyclic nucleotide-gated, and several other cation channels. The structure of one of its members, a K⁺ channel, reveals

cation-selective architectural features that are probably shared by all family members^{1,2}.

Much less is known about the mechanisms of ion selectivity in anion channels. The ClC Cl[−] channels, members of a large family of anion channels, are found throughout biology in both prokaryotic and eukaryotic cells^{3–5}. In many cases the biological roles of ClC Cl[−] channels are unknown, but in vertebrates several important functions are well established. In skeletal muscle, ClC Cl[−] channels stabilize the resting membrane potential and regulate electrical excitability⁶. In the kidney, ClC Cl[−] channels operate in concert with the Na⁺, K⁺, Cl[−] cotransporter and K⁺ channels to produce transepithelial fluid and electrolyte transport^{3,4}. Genetic defects of

Table 1 Summary of data collection and refinement statistics

Data collection	Beamline	Resolution (Å)	Completeness (%)	R_{merge}	$I/\sigma I$
StClC C2					
Native	BNL X25	35–3.5	99.7 (100)	5.2 (56.5)	20.0 (2.2)
Pt1	BNL X25	35–5.0	84.8 (81.3)	4.6 (19.0)	14.7 (3.9)
Pt2	BNL X25	35–4.2	99.3 (98.7)	6.1 (31.5)	14.8 (3.0)
Br [−]	BNL X25	35–4.5	86.4 (80.0)	5.7 (30.8)	15.2 (3.1)
EcClC P2 ₁ 2 ₁ 2 ₁					
Native	ESRF ID13	50–3.5	98.6 (91.3)	7.5 (50.5)	26.3 (1.8)
Se-Met	ESRF ID13	30–4.0	95.1 (97.5)	8.4 (40.5)	17.7 (5.7)
StClC P2 ₁					
Native	BNL X25	35–3.0	94.7 (96.6)	6.5 (54.4)	13.7 (1.7)
Phasing	R_{cullis} (acentric/centric)	Phasing power (acentric/centric)	No. of (sites)	Resolution (Å)	
Pt1	0.72/0.78	1.89/1.46	2	35–5.0	
Pt2	0.84/0.84	1.24/1.15	2	35–4.2	
Refinement	Asymmetric unit, no. of subunits	Resolution (Å)	$R_{\text{free}}/R_{\text{cryst}}$	r.m.s.d.* (bond/angles)	
<i>E. coli</i> P2 ₁ 2 ₁ 2 ₁	6	20–3.5	30.2/29.0	0.01/1.6	
<i>S. typhimurium</i> P2 ₁	4	20–3.0	28.8/25.5	0.01/1.5	

Data sets of the C2 ($a = 191$, $b = 99$, $c = 82$, $\alpha = \gamma = 90^\circ$, $\beta = 98.7^\circ$) and the P2₁ ($a = 185$, $b = 90$, $c = 81$, $\alpha = \gamma = 90^\circ$, $\beta = 98.9^\circ$) crystal form of StClC were collected at Brookhaven National Laboratory (BNL), station X25, with the Brandeis-CCD detector. Data sets of the P2₁2₁2₁ ($a = 105$, $b = 152$, $c = 263$, $\alpha = \beta = \gamma = 90^\circ$) crystal form of EcClC were collected at the European Synchrotron Radiation Facility (ESRF), station ID-13, with a MAR-CCD detector. Numbers in parentheses correspond to the highest resolution shell.

* Root mean square difference.

human CIC Cl⁻ channels underlie autosomal dominant and recessive forms of familial myotonia and several familial nephropathies including Bartter's syndrome and Dent's disease^{3,4,6}.

The study of CIC channels began about 20 years ago when Miller discovered a Cl⁻ channel from the electric organ of the torpedo ray (CIC-0)⁷. On the basis of quantitative single channel analysis, he predicted that these channels contain two parallel, independent pores (a double-barrel channel)^{8,9}. Jentsch cloned the gene for CIC-0, setting the stage for characterization of the entire CIC gene family¹⁰. Mutational, biochemical and electron microscopic structural analyses confirmed the presence of a double-barrel

channel¹¹⁻¹⁵. Many other properties of CIC Cl⁻ channels have been more difficult to discern in the absence of high-resolution structural data. For example, it has been difficult to define many aspects of the membrane topology, and the molecular components of the pore and selectivity filter have remained ambiguous. Different CIC Cl⁻ channels exhibit unique functional properties, particularly with respect to the signals that cause them to open and close (the process called gating), but they all seem to share two characteristics that are apparently inherent to this ion channel family. The first characteristic is anion over cation selectivity, manifest as permeation by the halogen ions Cl⁻ and Br⁻ (and in some cases I⁻), and

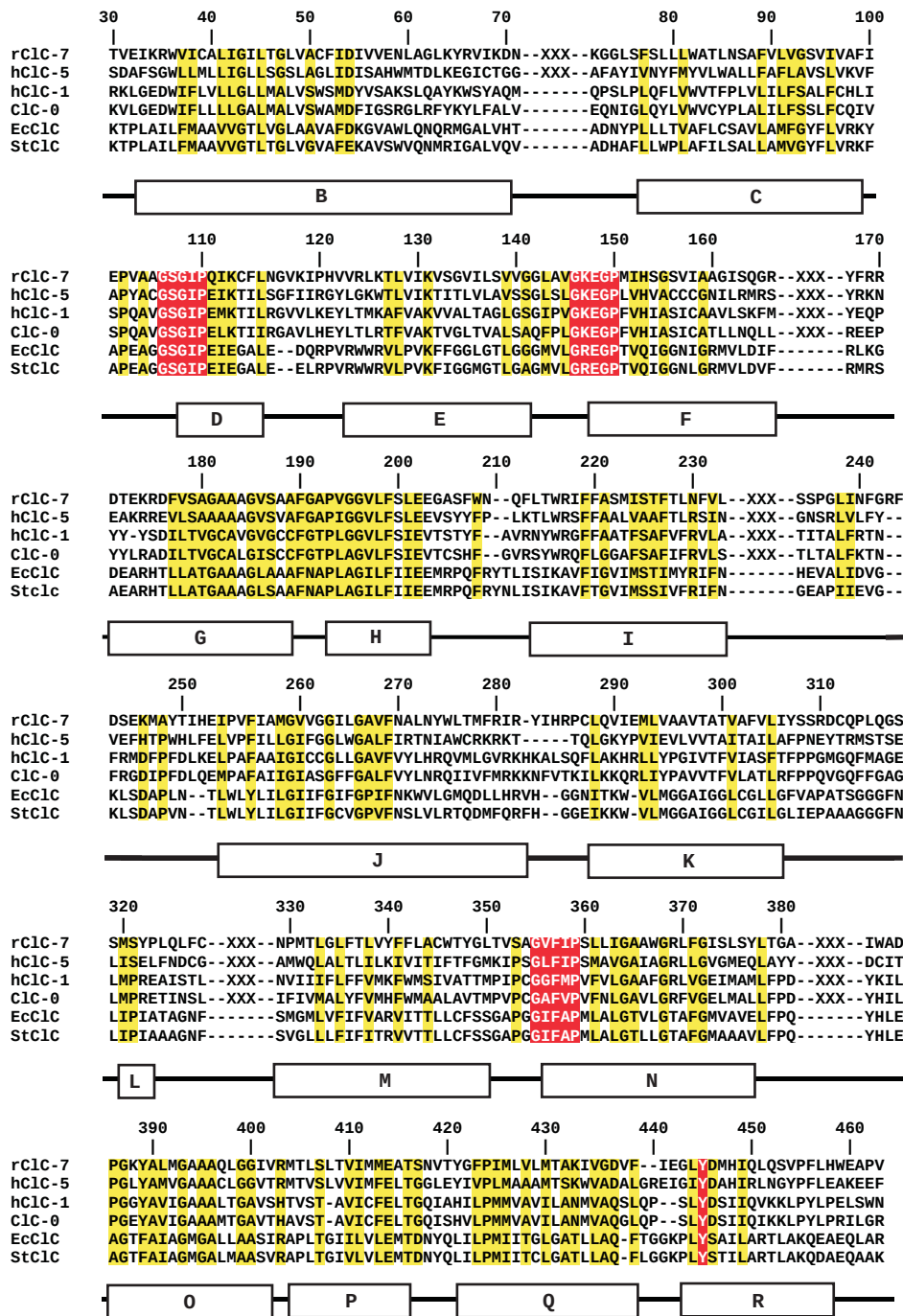


Figure 1 Sequence alignment of CIC channels. Secondary structure and numbering (*S. typhimurium*, StCIC channel) are indicated below and above the sequences, respectively. Residues from the Cl⁻ selectivity filter are red and regions of high homology are yellow. Insertions in eukaryotic CIC channels that are larger than ten residues (XXX) were cut out of the alignment. Amino-acid sequences are: rCIC-7, *Rattus norvegicus*

(GenBank number Z67743); hCIC-5, *Homo sapiens* (NM_000084); hCIC-1, *H. sapiens* (M97820); CIC-0, *Torpedo marmorata* (X56758); EcCIC, *E. coli* (NC_000913); StCIC, *S. typhimurium* (AE008704). The alignment was made with ClustalW⁴⁹ followed by manual adjustment.

block of Cl^- current by larger anions such as I^- , SCN^- and other organic anions^{16–18}. The second characteristic of CIC Cl^- channels is strong functional coupling between ion conduction and gating^{17,19–23}. The Cl^- ion is the current-carrying substrate that flows through the pore, and it is also an allosteric regulator of the channel's gates. Apparently by binding in the pore, the Cl^- ion influences the open probability, giving rise to properties such as voltage-dependent gating in some CIC Cl^- channels^{20,21}.

Amino-acid sequences of the integral membrane component of several CIC Cl^- channels are shown in Fig. 1, along with the secondary structure elements defined in the present study. All CIC Cl^- channels exhibit sequence conservation throughout, indicating

conservation of their three-dimensional structures. Prokaryotic CIC Cl^- channels depart from eukaryotic family members in the composition of their cytoplasmic amino and carboxy termini (terminal sequences not shown)²⁴. Little is known about the biology of prokaryotic CIC Cl^- channels. But given their sequence conservation, together with a recent study of ion selectivity in a CIC Cl^- channel from *E. coli*, we can be certain that a common set of ion selectivity principles applies to the entire family²⁵.

Structure

The structures of two CIC Cl^- channels from *S. typhimurium* (StCIC) and *E. coli* (EcCIC) were determined in parallel with data to 3.0 and 3.5 Å Bragg spacings, respectively (Table 1 and Methods). Initial phases were estimated from platinum derivatives of StCIC crystals. The building and refinement of accurate atomic models were aided by the identification of 17 methionine residues in selenium difference maps of the EcCIC channel (Fig. 2a) and by the presence of two-fold, four-fold and six-fold noncrystallographic symmetry (NCS) in a variety of crystal forms. The NCS was used to provide accurate phases for the calculation of electron density maps and to constrain the atomic coordinates throughout refinement²⁶ (Fig. 2b).

Two views of the StCIC Cl^- channel are shown in Fig. 3. The channel contains two identical subunits (red and blue), which are related by a two-fold axis of symmetry perpendicular to the membrane plane. Viewed along the two-fold axis from outside the cell, each subunit is triangular (Fig. 3a). The entire channel with two subunits is shaped like a rhombus, with major and minor diagonals of 100 and 55 Å, respectively. In the direction perpendicular to the membrane plane, the channel is about 65 Å thick owing to helical extensions that protrude into the aqueous solution (Fig. 3b). The contact surface area between subunits is extensive, encompassing nearly 2,300 Å² within the membrane. Such a large, stable interface is expected because CIC Cl^- channels are thought to exist and function only as dimers. The pore, or ion pathway, is not formed at the interface between subunits. Rather, each subunit forms its own independent pore and selectivity filter (green sphere). With regard to their pore stoichiometry, CIC Cl^- channels are like the porins²⁷, the aquaporins^{28,29} and bacteriorhodopsin³⁰, not like the cation channel family to which the K^+ channels belong. The former examples are trimers or tetramers of identical subunits in which each individual subunit forms an independent pore or active site. The K^+ channels, however, are tetramers in which the pore is formed by four identical subunits encircling a central ion pathway¹.

The structure of the EcCIC channel is essentially the same as the StCIC channel, except where deviations occur on the surface owing to crystal contacts. Likewise, deviations from perfect two-fold symmetry between the subunits within the dimeric channel probably result from protein contacts within the crystal. For example, the N terminus from only one of the subunits is visible as an α -helix near the cytoplasmic surface, whereas in the other subunit the N terminus is disordered (Fig. 3).

The CIC Cl^- channel subunit contains 18 α -helices (labelled A–R) and exhibits a complex topology (Fig. 4a). The three-dimensional structure reveals an internally repeated pattern in the sense that the N-terminal half of the polypeptide (α -helices B–I, green) is structurally related to the C-terminal half (J–Q, blue). The two similar structures have opposite orientations in the membrane, that is, they run antiparallel, and create a pseudo two-fold symmetry axis within the membrane. The internal symmetry is easily appreciated when viewing a subunit along the pseudo two-fold axis from the dimer interface (Fig. 4b). We refer to this arrangement as an antiparallel architecture. The internal repeat pattern in CIC Cl^- channels was not previously recognized on the basis of amino-acid sequence analysis, but careful alignment with knowledge of the protein structure shows that the two halves are indeed weakly correlated in their sequence, particularly with respect to

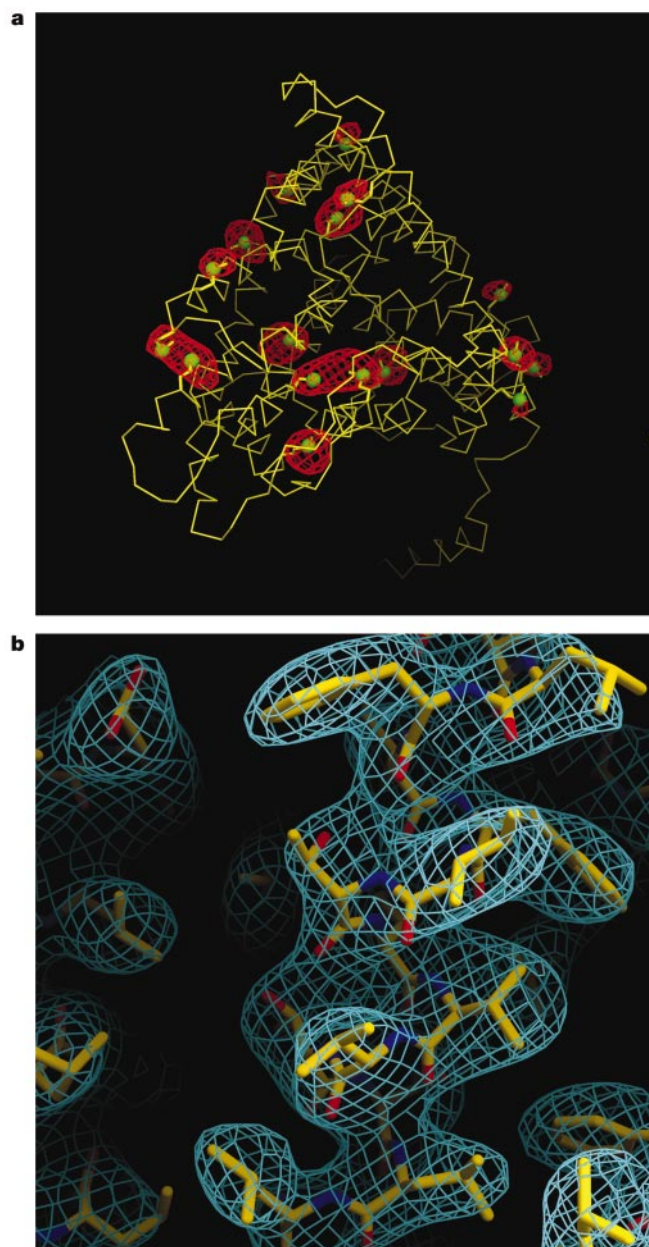


Figure 2 Experimental electron density. **a**, $F_o - F_{\text{native}}$ difference density (contoured at 4σ) in the EcCIC $P2_12_1$ crystal form superimposed on an α carbon trace of the EcCIC subunit viewed from outside the cell. Side chains of methionine residues are shown in stick representation with the position of the Se atom indicated by a green sphere. The difference Fourier map was calculated to 4.0 Å and averaged over the six subunits in the asymmetric unit. **b**, Electron density map from the StCIC $P2_1$ crystal form at 3.0 Å resolution, contoured at 1σ . The map was calculated from native amplitudes and solvent-flattened, averaged phases. The refined structure is shown as a stick model. Figures 2–6 were prepared with DINO (<http://www.dino3d.org>).

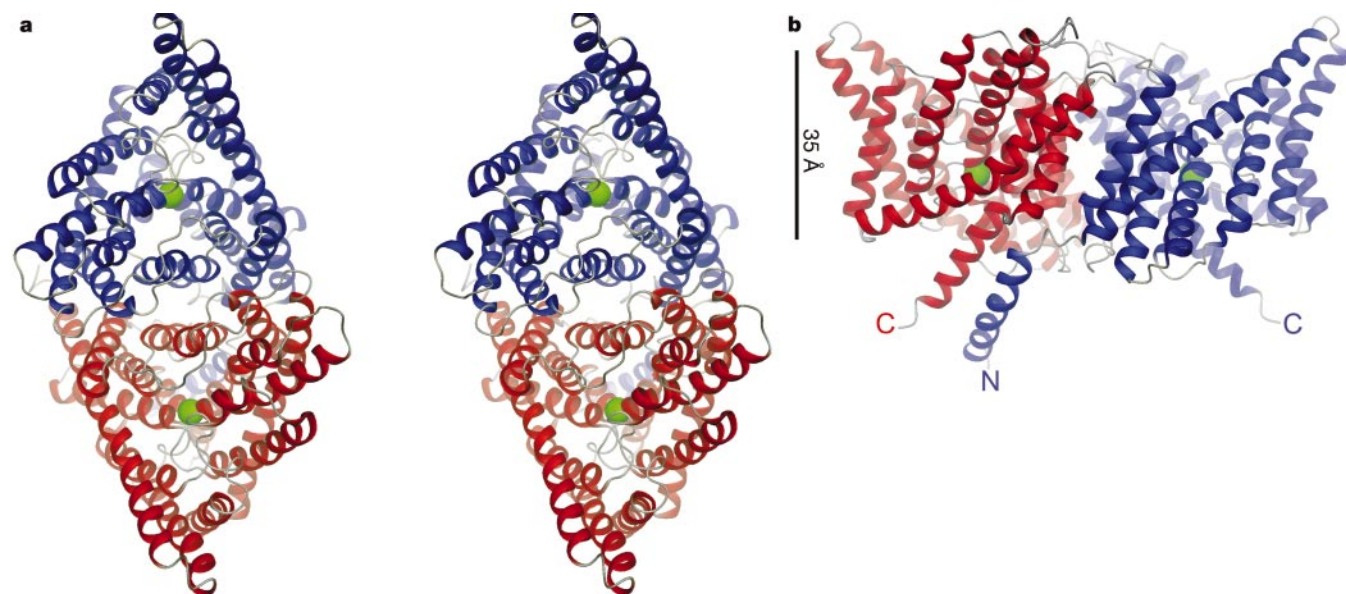


Figure 3 Structure of the StClC dimer. **a**, Stereo view of a ribbon representation of the StClC dimer from the extracellular side. The two subunits are blue and red. A Cl^- ion in the selectivity filter is represented as a green sphere. **b**, View from within the membrane with

the extracellular solution above. The channel is rotated by 90° about the x - and y -axes relative to **a**. The black line (35 Å) indicates the approximate thickness of the membrane. See Supplementary Information for the X-ray structure.

the disposition of glycine residues. It is interesting that an antiparallel architecture is also seen in the individual subunits of aquaporins (channels for water and small solute molecules), even though their membrane-spanning topology is unrelated to that of ClC Cl^- channels^{28,29}.

The transmembrane α -helices within a ClC subunit are remarkably tilted and variable in length (Fig. 4b). There is an important reason underlying this seemingly complicated arrangement of helices. The two halves of the subunit (Fig. 4b, green and blue) wrap around a common centre so as to bring together amino acids at the ends of α -helices from disparate segments of the polypeptide chain (Figs 1 and 4b, red). As we show below, these amino acids form the Cl^- ion selectivity filter within the membrane.

Anion selectivity

Important amino acids from four separate regions are brought together near the membrane centre to form an ion-binding site. These regions (red in the sequence alignments, Fig. 1) are highly conserved in ClC Cl^- channels; they include the sequences GSGIP (106–110), G(K/R)EGP (146–150) and GXFXP (355–359), as well as Tyr 445. It is significant that these sequences occur at the N termini of α -helices, where polypeptide loops precede α -helices

D, F and N (Figs 1 and 4a). As a result, each of these helices is oriented with its N terminus pointed towards the binding site. Because of the helix dipole, or N-terminal positive end charge, this arrangement of helices is expected to create an electrostatically favourable environment for anion binding (depicted by blue in Fig. 5a).

A strong peak of electron density is present in the ion-binding site in solvent-flattened, averaged experimental maps (not shown). To determine whether the peak is due to a bound Cl^- ion, we took advantage of the fact that ClC Cl^- channels conduct Br^- ions nearly as well as Cl^- ions. We grew crystals in a solution containing NaBr instead of NaCl. The experiment was carried out using a C2 crystal form (Table 1). Difference electron density corresponding to Fourier coefficients $F_{\text{Br}} - F_{\text{Cl}}$ (Fig. 5b, red) is shown superimposed on a $2F_o - F_c$ electron density map surrounding the corresponding position within the $P2_1$ crystal form (Fig. 5b, cyan). The strong positive difference peak is consistent with the replacement of a Cl^- ion by the more electron-dense Br^- ion. We therefore conclude that the ion present at this site is indeed a Cl^- ion. Because this site is formed by the most conserved amino acids in the ClC Cl^- channel family and contains a Cl^- ion, we conclude that it is the selectivity filter.

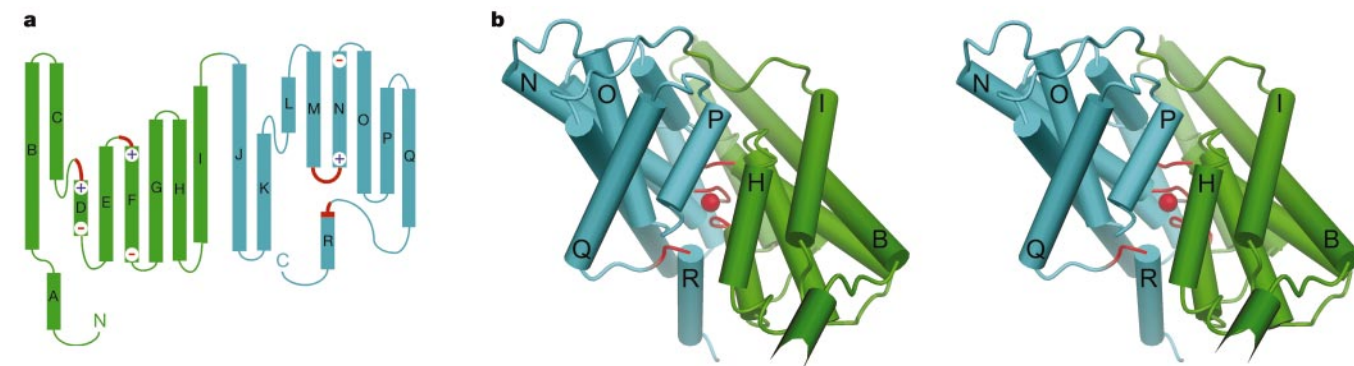


Figure 4 Structure of the StClC subunit. **a**, The α -helices (A–R) are drawn as cylinders with the extracellular region above and the intracellular region below. The two halves of the subunit are green and cyan, and regions forming the Cl^- selectivity filter are red. Partial charges at the end of helices involved in Cl^- binding are indicated by + and – (end

charges) to indicate the sense of the helix dipole. **b**, Stereo view of the StClC subunit viewed from within the plane of the membrane from the dimer interface with the extracellular solution above. The α -helices are drawn as cylinders, loop regions as cords (with the selectivity filter red), and the Cl^- ion as a red sphere.

The Cl^- ion is coordinated by main-chain amide nitrogen atoms from amino acids Ile 356 and Phe 357 and by the side-chain oxygen atoms from Ser 107 and Tyr 445 (Figs 1 and 5c). The nitrogen atoms are not involved in hydrogen bonding with the protein, and are available for ion binding because α -helix N begins just at position 356, allowing the 'free' pair of amide nitrogen atoms to form a cradle on one side of the Cl^- ion. The side chains of Ser 107 and Tyr 445 contact the Cl^- ion opposite the nitrogen cradle. The electrophilic aromatic ring makes the Tyr hydroxyl group a good proton donor and therefore an excellent Cl^- ion ligand. Likewise, the Ser 107 hydroxyl group is a good ligand because it caps the end of α -helix D. That is, by forming a hydrogen bond with the amide nitrogen of Ile 109, the serine hydroxyl is polarized to interact more strongly with the Cl^- ion. In addition to the interactions with polar functional groups, the Cl^- ion is surrounded by a number of hydrophobic amino-acid side chains.

It is noteworthy that in the Cl^- channel the ion does not make direct contact with a full positive charge from lysine or arginine residues. The favourable electrostatic environment for Cl^- arises instead from partial positive charges contributed by helix dipole interactions and by contacts with main-chain and side-chain nitrogen and oxygen atoms. We suggest that a full positive charge would create a deep energy well and cause a Cl^- ion to bind too tightly. Thus, it would appear that evolution of the channel has

resulted in partial charges to stabilize a Cl^- ion and still permit rapid ionic diffusion rates.

Conduction pore

The Cl^- ion conduction pore is shown in Fig. 6a, viewed from within the membrane with the extracellular solution above and the intracellular solution below. The picture was made by cutting the subunit in half along a plane that is perpendicular to the pseudo two-fold axis. The N termini of pseudo symmetry-related α -helices F and N, together with that of D, point at a narrow tunnel, about 12 Å long, towards the centre of the membrane. Wider, water-filled vestibules reach in from the extracellular and intracellular solutions (Fig. 6a, blue mesh). The Cl^- ion is shown at its position in the pore, about 2 Å below the pseudo two-fold axis. Unexpectedly, just above the pseudo two-fold axis, a glutamate side chain projects into the pore. Glu 148, a rather conserved residue in ClC Cl^- channels, is sandwiched between the positive ends of α -helices F and N, hydrogen bonded to its own amide nitrogen atom. The selectivity filter effectively has two anions in it, a Cl^- ion closer to the intracellular vestibule (inner site) and a carboxylate anion closer to the extracellular vestibule (outer site). For Cl^- conduction to occur, the glutamate side chain presumably would have to swing out of the way (perhaps associated with structural changes extending beyond the glutamate side chain), permitting the entry of a Cl^- ion in its place.

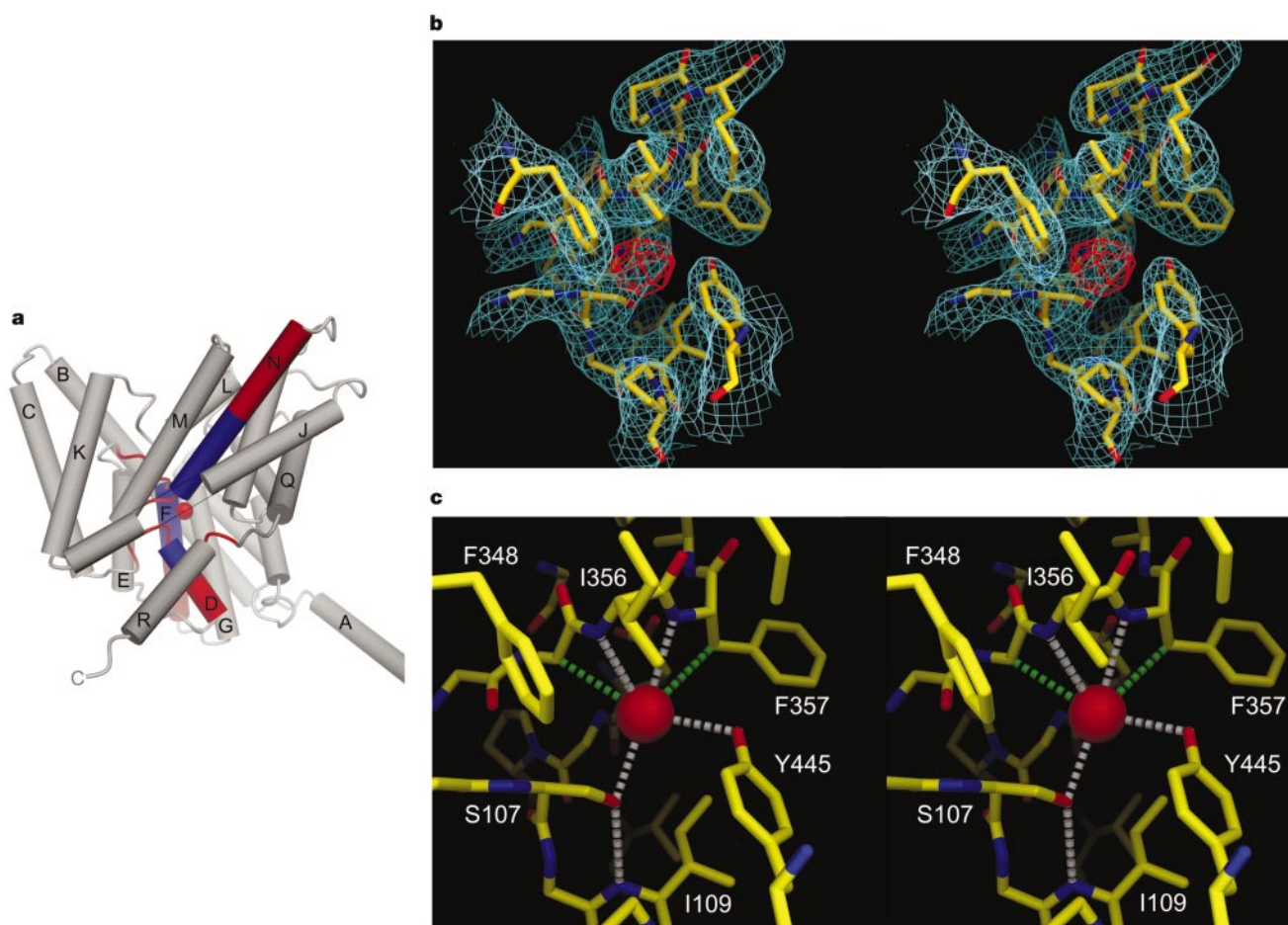


Figure 5 Structure of the StClC selectivity filter. **a**, Helix dipoles (end charges) point towards the selectivity filter. The α -helices are shown as cylinders with labels corresponding to Fig. 4a. The amino (positive, blue) and carboxy (negative, red) ends of α -helices D, F and N are shown. The selectivity filter residues are shown as red cords surrounding a Cl^- ion (red sphere). The view is from 20° below the membrane plane; the dimer interface is to the right, and the extracellular solution above. Part of α -helix J has been removed for clarity (grey line). **b**, Stereo view of electron density in the selectivity

filter. A $2F_o - F_c$ Cl^- omit map (StClC P2₁ crystal form, 3.0 Å resolution, 1 σ contour, cyan) is superimposed on a stick model of residues selected within a 5 Å radius of the Cl^- ion. An $F_{\text{Br}} - F_{\text{Cl}}$ difference Fourier map (StClC C2 crystal form, 4.5 Å resolution, 5 σ contour, red) is shown transformed to the corresponding position in the P2₁ unit cell. **c**, Stereo view of the Cl^- ion-binding site. Distances (<3.6 Å) to the Cl^- ion (red sphere) are shown for polar (white dashed lines) and hydrophobic (green dashed lines) contacts. A hydrogen bond between Ser 107 and the amide nitrogen of Ile 109 is shown (white dashed line).

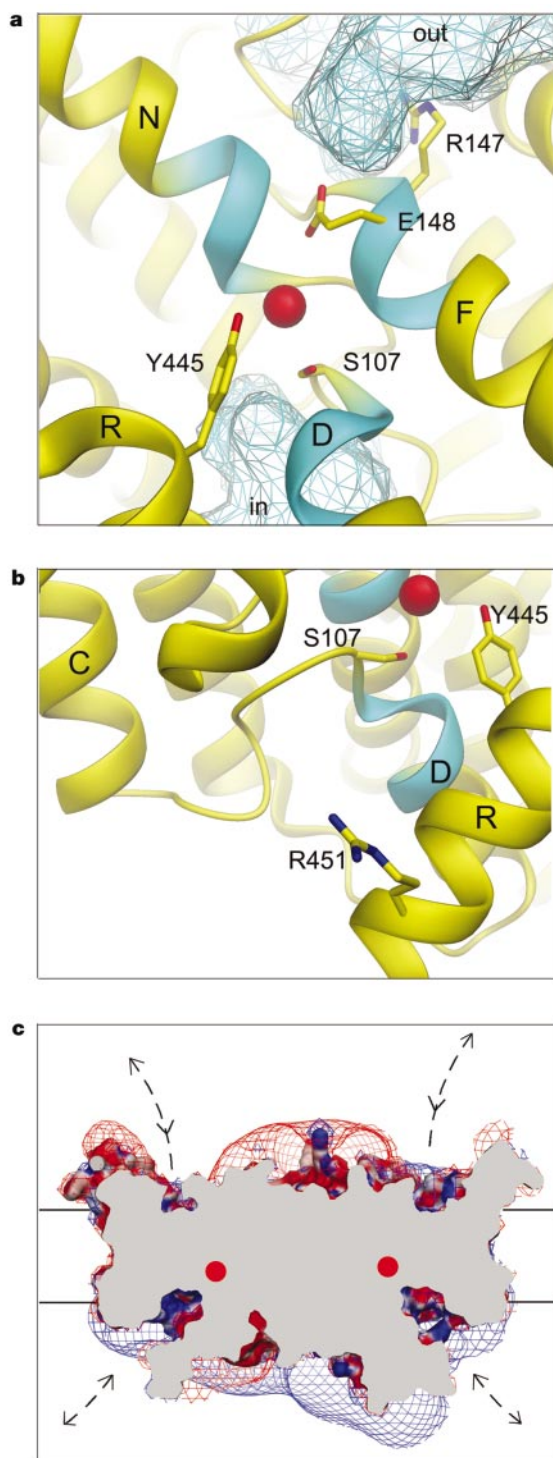


Figure 6 Ion conduction pathway. **a**, The ion-binding site viewed from the dimer interface, along the pseudo two-fold axis, with foreground α -helices removed for clarity. The protein is shown as a ribbon with selected residues as sticks. The amino terminal ends of α -helices D, F and N are cyan and Cl^- is shown as a red sphere. Aqueous cavities approaching the selectivity filter from the extracellular solution (out) and intracellular solution (in) are shown as a cyan mesh. **b**, View of the intracellular channel entrance. The molecule is rotated by about 90° around the y -axis relative to **a**. **c**, Surface electrostatic potential on the ClC dimer in 150 mM electrolyte. The channel is sliced in half to show the pore entryways (but not the full extent of their depth) on the extracellular (above) and intracellular (below) sides of the membrane. Isocontour surfaces of -12 mV (red mesh) and $+12$ mV (blue mesh) are shown. Cl^- ions are shown as red spheres. Dashed lines highlight the pore entryways. The solvent-excluded surface of the StClC dimer was calculated with the program MSMS⁵⁰.

The structure of this narrow, anion-selective region of the channel is interesting in light of biophysical studies on the influence of extracellular Cl^- ion concentration on the function of ClC Cl^- channels. Some ClC channels have been described as being Cl^- -activated chloride channels, because a certain amount of Cl^- is required in the extracellular solution to gate the pore open²⁰. Moreover, Cl^- appears to exert its gating effect by actually entering the pore, meaning that gating (the process of opening the pore) and ion conduction (the process of ions moving through the open pore) are closely linked^{20,21}. An additional important piece of information comes from the observation, made in ClC-1 Cl^- channels, that a series of anions rank differently in their effects on conduction and gating¹⁷. To explain these findings, two anion-binding sites in the pore have been hypothesized, one accounting for 'conduction selectivity' (proposed to be closer to the intracellular solution) and the other for 'gating selectivity' (proposed to be closer to the extracellular solution). The structure offers a possible explanation for these findings. The inner site, where a Cl^- ion is specifically coordinated, probably is important for conduction selectivity as ions diffuse through the open pore. What we are calling the outer site would prevent the channel from conducting as long as the glutamate side chain 'blocks' the ion pathway. However, if Cl^- ions from the extracellular solution can enter the pore and induce a conformational change that displaces the glutamate side chain (that is, by competing with the carboxylate group), then the property of Cl^- activation could be explained. Such a mechanism would imply that when the pore is open, conduction occurs through the interaction of two closely spaced Cl^- ions in the selectivity filter.

The vestibules leading up to the selectivity filter on both sides of the membrane contain basic (positively charged) amino acids, for example Arg 147 (Fig. 6a) and Arg 451 (Fig. 6b). The distribution of charges on the entire channel surface creates an electrostatic potential that probably funnels Cl^- ions into the pore entryways (Fig. 6c). The two pores on the dimer are separated by a large distance and by an electronegative (Cl^- -repulsive) region on the extracellular surface (Fig. 6c, red). These features are consistent with the functional independence of the two pores in ClC-0 channels⁸.

The effects of numerous mutations on the properties of ClC-0 and ClC-1 Cl^- channels have been studied^{3–5}. We mention only a few of these results as they relate to the focus of this paper, the mechanism of anion selectivity. Mutation of serine to threonine at the Cl^- binding site in ClC-0 (Ser 107 of StClC, Figs 1 and 5c) affected selectivity among anions and dramatically reduced conductance¹¹. Mutations near the extracellular entryway (StClC 146–148; Figs 1 and 6a) and intracellular entryway (StClC α -helix R, Figs 4b and 6b) of ClC-0 also altered ion conduction as well as gating^{12,20,23,31,32}. One site in particular, Lys 519 of ClC-0 (adjacent to Arg 451 of StClC, Fig. 6b) appears to influence conduction mainly by an electrostatic interaction, whereby the positively charged amino group attracts Cl^- ions and increases the conductance^{12,23}. The results of these mutational studies are in good agreement with the X-ray structure.

The location of α -helix R in the structure is intriguing (Figs 4b and 6b). Its N-terminal end begins at the selectivity filter, where Tyr 445 interacts directly with the Cl^- ion, and its C-terminal end projects into the cytoplasm. As a point of speculation, α -helix R would seem to offer a direct route for regulating channel function through processes occurring in the cytoplasm.

Discussion

The ClC Cl^- channel is a homodimer with a two-fold axis perpendicular to the membrane plane; each of the subunits within the dimer forms its own ion-conduction pore. Each individual subunit, or channel-forming unit, exhibits an antiparallel architecture (Fig. 7a). Why did such a structural plan evolve? We propose that the antiparallel architecture satisfies two fundamental physical

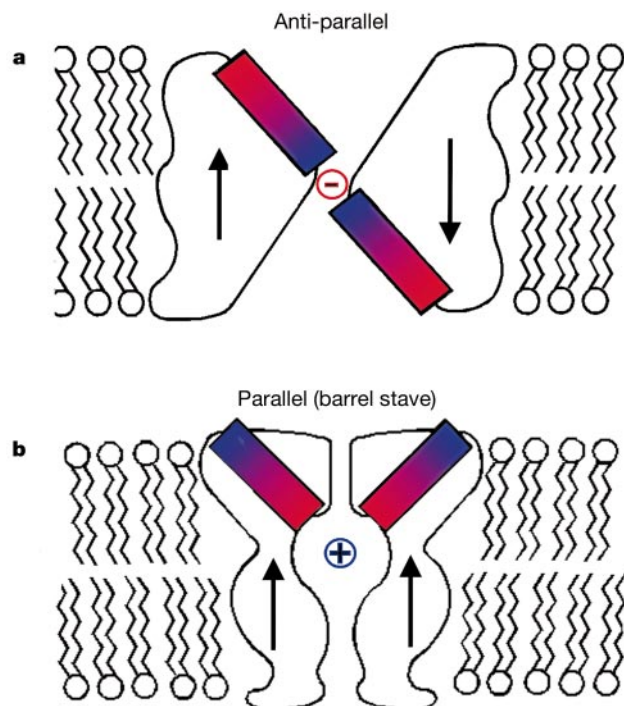


Figure 7 Two architectures for ion-channel proteins. **a**, The antiparallel architecture of ClC Cl^- channels contains structurally similar halves with opposite orientations in the membrane (arrows). This architecture permits like ends (same dipole sense) of α -helices to point at the membrane centre from opposite sides of the membrane (180° separation). **b**, The parallel or barrel stave architecture of K^+ channels contains structurally similar or identical subunits with the same membrane orientation (arrows). Helices point at the membrane centre from the same side of the membrane. Helices are depicted as dipoles with blue (positive) and red (negative) ends.

constraints on membrane-transport proteins. First, active sites or selectivity filters within the membrane exploit the polar ends of α -helices to stabilize (or destabilize) substrates or ions through electrostatic and chemical interactions. Second, the opposite ends of the helices (away from the active site) have to be directed into the aqueous environment outside the membrane. This second constraint stems from the simple fact that it is energetically unfavourable to bury the polar end of a helix in the low-dielectric membrane core. An antiparallel architecture satisfies these constraints elegantly, allowing like helix ends (same dipole orientation) to point at the selectivity filter from opposite directions (with a 180° angular separation, Figs 6a and 7a). There are other ways to satisfy the above constraints on membrane helices and active sites, as in the case of the K^+ channel (Fig. 7b), but the antiparallel architecture is a fascinating solution to the problem.

Anion selectivity is achieved in ClC Cl^- channels through partial positive charges, not through full positive charges. This principle of partial charges is also used in the K^+ channel, but with reversed polarity, in the form of negative helix dipoles and main-chain carbonyl oxygen atoms^{1,2}. Partial charges create a favourable electrostatic environment for relatively large monovalent ions such as Cl^- (radius 1.81 Å) or K^+ (radius 1.33 Å) while preventing the ions from binding too tightly—a requirement for high conduction rates. In the ClC Cl^- channels, partial charges are contributed by positive helix dipoles, by main-chain amide nitrogen atoms, and by side-chain hydroxyl groups of serine and tyrosine residues. The amino acids forming the selectivity filter are quite conserved in ClC Cl^- channels, implying that the anion-selective principles described here apply to the entire family.

We wondered about the striking architectural difference between Cl^- and K^+ channels: the Cl^- channel has a roughly hourglass-

shaped pore with a narrow constriction (its selectivity filter) near the membrane centre (Fig. 7a), whereas the K^+ channel has a widening, or cavity, at the membrane centre (Fig. 7b). The cavity in K^+ channels, working in concert with helix dipoles, helps to overcome the dielectric barrier, or energy barrier experienced by an ion owing to the low-dielectric membrane environment^{1,33}. For some reason the Cl^- channel uses only helix dipoles. Is the membrane dielectric barrier different for K^+ and Cl^- ions? We do not know the full answer, but we do know of one difference that might make the dielectric barrier smaller for anions. Biological membranes have a dipole layer on the plane of the water-lipid interface that contributes a positive electrostatic potential, perhaps as much as +300 mV (ref. 34), to the membrane interior^{35,36}. The precise origin of this dipole layer is unknown, but its presence accounts for the fact that hydrophobic anions partition into biological membranes with greater ease than hydrophobic cations³⁴. In other words, all else being equal, anions are more stable than cations inside the membrane. Perhaps for this reason Cl^- channels do not require a wide, water-filled pore near the membrane centre to minimize the dielectric barrier.

ClC Cl^- channels are complex; further work will be required to understand how they gate and how the gating processes are coupled to ion conduction through the pore. We have provided here a structural basis for guiding further work on ClC Cl^- channel function, and established the physical and chemical principles underlying ion selectivity in an anion channel. □

Methods

Protein preparation

EcClC (GenBank accession number NC_000913) and StClC (AE008704) were cloned into a pET28b+ vector using *Nco*I and *Xho*I restriction sites and protein expressed with a C-terminal histidine tag in *E. coli* strain BL21 DE3. In StClC, Met 26 was mutated to Leu and Cys 264 was mutated to Val. Homogeneous expression of the full-length protein was confirmed by matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS). Bacteria were grown to an absorbance at 600 nm (A_{600}) of 1.0 in the presence of 50 mg l⁻¹ kanamycin at 37 °C, and induced with 200 μM isopropyl- β -D-thiogalactopyranoside (IPTG) overnight at 25 °C. Cells were centrifuged, lysed by French press or sonication in 50 mM Tris-HCl buffer at pH 7.5, 150 mM NaCl, 200 mg l⁻¹ lysozyme, 20 mg l⁻¹ DNase, leupeptin and pepstatin, and 1.0 mM phenylmethyl sulphonyl fluoride (PMSF), and extracted with 50 mM *n*-decyl- β -maltoside (Anatrace, Sol-Grade D322S) for 2 h at room temperature. Extract was centrifuged for 40 min at 30,000g and the supernatant loaded onto a cobalt-affinity column (Talon Metal Affinity Resin, Clontech). The column was washed with imidazole (gradient 0–40 mM) and desired protein eluted with 400 mM imidazole. Protein was concentrated to 15–20 mg ml⁻¹, the His tag was removed by proteolytic digestion with 5 international units (IU) thrombin (StClC) or 1.5 IU endoprotease Lys-C (EcClC) for 2 h at room temperature, and protein was then run on a Superdex 200 column in 45 mM *n*-octyl- β -maltoside, 150 mM NaCl and 10 mM Tris-HCl at pH 7.5. Pooled peaks were concentrated to 10–15 mg ml⁻¹ and dialysed overnight against 45 mM *n*-octyl- β -maltoside, 75 mM NaCl, 10 mM Tris-HCl at pH 7.5. Protein for Br⁻ soaks were dialysed against 45 mM *n*-octyl- β -maltoside, 75 mM NaBr and 10 mM Tris-sulphate buffer at pH 7.5. For protein labelled with seleno-L-methionine, transformed bacteria (B834 DE3) were grown in minimal medium (L-methionine replaced by seleno-L-methionine), and the protein was purified as above. Substitution of methionine sulphur with selenium was confirmed by MALDI-MS.

Crystal preparation

Crystals of EcClC and StClC channels were grown in sitting drops at 20 °C by equilibrating a 1:1 mixture of protein and reservoir solutions against the reservoir. For the EcClC $P_{21}2_12_1$ crystal form (with and without selenium) the reservoir solution contained 31–34% polyethylene glycol (PEG 400), 50 mM Tris at pH 8.5, 50 mM Na₂SO₄ and 50 mM Li₂SO₄. The crystals were frozen in their mother liquor in the cryogenic nitrogen stream. The C2 and the P_{21} crystal forms of StClC crystallized from the same reservoir solution, which contained 26–31% PEG 400, 50 mM sodium acetate at pH 4.6, 75–100 mM Na₂SO₄ and 75–100 mM Li₂SO₄. PEG 400 was increased to 37% by dialysis before freezing in the nitrogen stream. For derivatization, crystals were soaked for 6 h in mother liquor containing 1 mM Pt(II)-terpyridine chloride (Aldrich).

Crystallography

All data sets were collected on frozen crystals on beam-lines BNL X25 and ESRF ID-19 (Table 1). The data were indexed and integrated with DENZO and SCALEPACK³⁷ and further processed with the CCP4 programs³⁸. Anomalous data for the Pt derivatives were collected at a wavelength of 1.07 Å. Heavy-metal positions for the C2 crystal form of StClC were identified by analysis of difference Patterson maps and by difference Fourier methods. Heavy-metal parameters were refined at 5 Å with SHARP³⁹. SOLOMON was used for iterative solvent flattening and phase combination⁴⁰. NCS operators were

estimated with GETAX⁴¹. Two-fold averaging and phase extension to 4.5 Å with RAVE⁴² allowed the building of a poly-alanine helix model, which served as a search model for molecular replacement into the P₂,2,2, crystal form of EcCIC using AMORE⁴³. Phases in the P₂,2,2, crystal form were extended to a resolution of 3.5 Å by six-fold NCS averaging with RAVE⁴². A model consisting of residues 12–461 was built into the averaged density using program O⁴⁴. The locations of 17 methionine residues were identified by peaks ($>4\sigma$) in a $F_{\text{Se}} - F_{\text{native}}$ difference Fourier map. Refinement was carried out in several simulated annealing cycles using the maximum likelihood target implemented in CNS⁴⁵ followed by manual inspection and rebuilding of the molecule. Six-fold NCS constraints were used throughout refinement. Restrained individual *B*-factor refinement was justified by favourable observable-to-parameter ratio and a further drop of R_{free} (30.2%) and R_{cryst} (29.0%). The refined EcCIC dimer was used as search model for molecular replacement into a P₂ crystal form of StCIC⁴³. Four-fold NCS averaging and phase extension to 3.0 Å using RAVE⁴² and DM⁴⁶ was followed by rebuilding of the model into the averaged electron density⁴⁴. Initial cycles of simulated annealing refinement were carried out maintaining four-fold NCS constraints⁴⁵. In later stages, the strict constraints were loosened within the two subunits of the dimer. The α -helix H–I loop, which differed in conformation between the two subunits of the same dimer, was rebuilt and the N terminus of subunit B was extended by 29 residues added onto α -helix A. These residues were not visible in subunit A. In further refinement cycles, the strict NCS constraints were maintained only between the two dimers in the asymmetric unit. Restrained, individual *B*-factors were refined. The final model has agreement factors R_{free} and R_{cryst} of 28.8% and 25.5%, respectively, and contains 6,574 protein atoms, two Cl[−] ions, two sulphate ions (on the intracellular surface not near the pore), two lipids and one water. The presence of a halogen ion in the Cl[−]-binding site was confirmed with an $F_{\text{Br}} - F_{\text{Cl}}$ difference Fourier calculation in the StCIC C2 crystal form.

Poisson–Boltzmann calculation

Electrostatic potential was calculated with DELPHI⁴⁷. The CIC dimer, assigned a dielectric constant $\epsilon = 2$, was embedded in a membrane slab 18 Å thick ($\epsilon = 2$) and surrounded by water ($\epsilon = 80$, ionic strength 150 mM uni-univalent electrolyte). Partial protein charges were derived from the CHARMM 22 all-hydrogen force field⁴⁸; hydrogen positions were generated with the program CNS⁴⁵. Standard charging (neutral pH) was assumed for ionizable residues except Glu 235, which forms a carboxyl-carboxylate and therefore was uncharged. Electrostatic potential was calculated for a $99 \times 99 \times 99$ Å volume (grid spacing 0.8 Å).

Received 23 November; accepted 17 December 2001.

- Doyle, D. A. *et al.* The structure of the potassium channel: molecular basis of K⁺ conduction and selectivity. *Science* **280**, 69–77 (1998).
- Zhou, Y., Cabral, J. M., Kaufman, A. & MacKinnon, R. Chemistry of ion coordination and hydration revealed by a K⁺ channel–Fab complex at 2.0 Å resolution. *Nature* **414**, 43–48 (2001).
- Jentsch, T. J., Friedrich, T., Schriever, A. & Yamada, H. The CLC chloride channel family. *Pflügers Arch.* **437**, 783–795 (1999).
- Maduke, M., Miller, C. & Mindell, J. A. A decade of CLC chloride channels: structure, mechanism, and many unsettled questions. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 411–438 (2000).
- Fahlke, C. Ion permeation and selectivity in CLC-type chloride channels. *Am. J. Physiol. Renal Physiol.* **280**, F748–F757 (2001).
- Steinmeyer, K., Lorenz, C., Pusch, M., Koch, M. C. & Jentsch, T. J. Multimeric structure of CLC-1 chloride channel revealed by mutations in dominant myotonia congenita (Thomsen). *EMBO J.* **13**, 737–743 (1994).
- White, M. M. & Miller, C. A voltage-gated anion channel from the electric organ of *Torpedo californica*. *J. Biol. Chem.* **254**, 10161–10166 (1979).
- Miller, C. Open-state substructure of single chloride channels from *Torpedo electroplax*. *Phil. Trans. R. Soc. Lond. B* **299**, 401–411 (1982).
- Miller, C. & White, M. M. Dimeric structure of single chloride channels from *Torpedo electroplax*. *Proc. Natl Acad. Sci. USA* **81**, 2772–2775 (1984).
- Jentsch, T. J., Steinmeyer, K. & Schwarz, G. Primary structure of *Torpedo marmorata* chloride channel isolated by expression cloning in *Xenopus* oocytes. *Nature* **348**, 510–514 (1990).
- Ludewig, U., Pusch, M. & Jentsch, T. J. Two physically distinct pores in the dimeric CLC-0 chloride channel. *Nature* **383**, 340–343 (1996).
- Middleton, R. E., Pheasant, D. J. & Miller, C. Homodimeric architecture of a CLC-type chloride ion channel. *Nature* **383**, 337–340 (1996).
- Weinreich, F. & Jentsch, T. J. Pores formed by single subunits in mixed dimers of different CLC chloride channels. *J. Biol. Chem.* **276**, 2347–2353 (2001).
- Maduke, M., Pheasant, D. J. & Miller, C. High-level expression, functional reconstitution, and quaternary structure of a prokaryotic CLC-type chloride channel. *J. Gen. Physiol.* **114**, 713–722 (1999).
- Mindell, J. A., Maduke, M., Miller, C. & Grigorieff, N. Projection structure of a CLC-type chloride channel at 6.5 Å resolution. *Nature* **409**, 219–223 (2001).
- White, M. M. & Miller, C. Probes of the conduction process of a voltage-gated Cl[−] channel from *Torpedo electroplax*. *J. Gen. Physiol.* **78**, 1–18 (1981).
- Rychkov, G. Y., Pusch, M., Roberts, M. L., Jentsch, T. J. & Bretag, A. H. Permeation and block of the skeletal muscle chloride channel, CLC-1, by foreign anions. *J. Gen. Physiol.* **111**, 653–665 (1998).
- Rychkov, G., Pusch, M., Roberts, M. & Bretag, A. Interaction of hydrophobic anions with the rat skeletal muscle chloride channel CLC-1: effects on permeation and gating. *J. Physiol.* **530**, 379–393 (2001).
- Richard, E. A. & Miller, C. Steady-state coupling of ion-channel conformations to a transmembrane ion gradient. *Science* **247**, 1208–1210 (1990).
- Pusch, M., Ludewig, U., Rehfeldt, A. & Jentsch, T. J. Gating of the voltage-dependent chloride channel CLC-0 by the permeant anion. *Nature* **373**, 527–531 (1995).
- Chen, T. Y. & Miller, C. Nonequilibrium gating and voltage dependence of the CLC-0 Cl[−] channel. *J. Gen. Physiol.* **108**, 237–250 (1996).
- Ludewig, U., Jentsch, T. J. & Pusch, M. Inward rectification in CLC-0 chloride channels caused by mutations in several protein regions. *J. Gen. Physiol.* **110**, 165–171 (1997).

- Ludewig, U., Jentsch, T. J. & Pusch, M. Analysis of a protein region involved in permeation and gating of the voltage-gated *Torpedo* chloride channel CLC-0. *J. Physiol.* **498** (part 3), 691–702 (1997).
- Maduke, M., Williams, C. & Miller, C. Formation of CLC-0 chloride channels from separated transmembrane and cytoplasmic domains. *Biochemistry* **37**, 1315–1321 (1998).
- Maduke, M., Pheasant, D. J. & Miller, C. High-level expression, functional reconstitution, and quaternary structure of a prokaryotic CLC-type chloride channel. *J. Gen. Physiol.* **114**, 713–722 (1999).
- Kleywegt, G. J. & Read, R. J. Not your average density. *Structure* **5**, 1557–1569 (1997).
- Cowan, S. W. *et al.* Crystal structures explain functional properties of two *E. coli* porins. *Nature* **358**, 727–733 (1992).
- Murata, K. *et al.* Structural determinants of water permeation through aquaporin-1. *Nature* **407**, 599–605 (2000).
- Fu, D. *et al.* Structure of a glycerol-conducting channel and the basis for its selectivity. *Science* **290**, 481–486 (2000).
- Grigorieff, N., Ceska, T. A., Downing, K. H., Baldwin, J. M. & Henderson, R. Electron-crystallographic refinement of the structure of bacteriorhodopsin. *J. Mol. Biol.* **259**, 393–421 (1996).
- Fahlke, C., Yu, H. T., Beck, C. L., Rhodes, T. H. & George, A. L. Jr Pore-forming segments in voltage-gated chloride channels. *Nature* **390**, 529–532 (1997).
- Fahlke, C., Beck, C. L. & George, A. L. Jr A mutation in autosomal dominant myotonia congenita affects pore properties of the muscle chloride channel. *Proc. Natl Acad. Sci. USA* **94**, 2729–2734 (1997).
- Roux, B. & MacKinnon, R. The cavity and pore helices in the KcsA K⁺ channel: electrostatic stabilization of monovalent cations. *Science* **285**, 100–102 (1999).
- Andersen, O. S. & Fuchs, M. Potential energy barriers to ion transport within lipid bilayers. *Biophys. J.* **15**, 795–830 (1975).
- McLaughlin, S. Electrostatic potentials at membrane-solution interfaces. *Curr. Top. Membr. Transp.* **9**, 71–144 (1977).
- Jordan, P. C. Electrostatic modeling of ion pores. II. Effects attributable to the membrane dipole potential. *Biophys. J.* **41**, 189–195 (1983).
- Otwiński, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Collaborative Computational Project No. 4. The CCP4 Suite: Programs for X-ray crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- de La Fortelle, E. & Briconne, G. in *Methods in Enzymology* (eds Carter, C. W. & Sweet, R. M.) 472–494 (Academic, New York, 1997).
- Abrahams, J. P. & Leslie, A. G. W. Methods used in the structure determination of bovine mitochondrial F₁ ATPase. *Acta Crystallogr. D* **52**, 30–42 (1996).
- Vonrhein, C. & Schulz, G. E. Locating proper non-crystallographic symmetry in low-resolution electron-density maps with the program GETAX. *Acta Crystallogr. D. Biol. Crystallogr.* **55** (part 1), 225–229 (1999).
- Kleywegt, G. J. & Jones, T. A. in *Proc. CCP4 Study Weekend* (eds Bailey, S., Hubbard, R. & Waller, D.) 59–66 (Daresbury Laboratory, Daresbury, UK, 1994).
- Navaza, J. AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A* **50**, 157–163 (1994).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Brunker, A. T. *et al.* Crystallography & NMR system: new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Cowan, K. DM: an automated procedure for phase improvement by density modification. *Joint CCP4 ESF-EACBM Newsletter. Protein Crystallogr.* **31**, 34–38 (1994).
- Rocchia, W. *et al.* Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: Application to the molecular systems and geometric objects. *J. Comp. Chem.* **23**, 128–137 (2002).
- MacKerell, A. D. J. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
- Sanner, M. F., Olson, A. J. & Spehner, J. C. Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers* **38**, 305–320 (1996).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com>).

Acknowledgements

We thank M. Becker and L. Berman at the National Synchrotron Light Source X-25, the staff at the Cornell High Energy Synchrotron Source A1 and F1, and C. Petosa and A. Perrakis of the European Molecular Biology Laboratory at the European Synchrotron Radiation Facility ID-13 for assistance in data collection; J. Bonanno for technical assistance; members of the MacKinnon laboratory for assistance at all stages of the project; and S. Harrison, D. Jeruzalmi, A. Philippsen and O. Andersen for discussions. This work was supported by grants from the National Institutes of Health (NIH) to R.M. and from the National Center for Research Resources, NIH, to B.T.C. R.M. is an investigator in the Howard Hughes Medical Institute.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.M. (e-mail: mackinn@rockvax.rockefeller.edu). Coordinates have been deposited with the Protein Data Bank under accession codes 1KPK and 1KPL.

Detection of carbonates in dust shells around evolved stars

F. Kemper*, C. Jäger†, L. B. F. M. Waters*‡, Th. Henning†, F. J. Molster§||, M. J. Barlow¶, T. Lim# & A. de Koter*

* Astronomical Institute "Anton Pannekoek", University of Amsterdam, Kruislaan 403, 1098 SJ Amsterdam, The Netherlands

† University of Jena, Astrophysical Institute and University Observatory (AIU), Schillergässchen 3, D-07745 Jena, Germany

‡ Instituut voor Sterrenkunde, KU Leuven, Celestijnenlaan 200B, 3001 Heverlee, Belgium

§ School of Materials Science and Engineering, Georgia Tech, Atlanta, Georgia 30332-0245, USA

|| ESTEC/ESA, Solar System division, Keplerlaan 1, 2201 AZ Noordwijk, The Netherlands

¶ Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK

Space Science Department, Rutherford Appleton Laboratory, Chilton, Didcot OX11 0QX, UK

Carbonates on large Solar System bodies like Earth and Mars^{1,2} (the latter represented by the meteorite ALH84001) form through the weathering of silicates in a watery (CO₃)²⁻ solution. The presence of carbonates in interplanetary dust particles and asteroids (again, represented by meteorites) is not completely understood, but has been attributed to aqueous alteration on a large parent body, which was subsequently shattered into smaller pieces. Despite efforts³⁻⁵, the presence of carbonates outside the Solar System has hitherto not been established^{6,7}. Here we report the discovery of the carbonates calcite and dolomite in the dust shells of evolved stars, where the conditions are too primitive for the formation of large parent bodies with liquid water. These carbonates, therefore, are not formed by aqueous alteration, but perhaps through processes on the surfaces of dust or ice grains or gas phase condensation. The presence of carbonates which did not form by aqueous alteration suggests that some of the carbonates found in Solar System bodies no longer provide direct evidence that liquid water was present on large parent bodies early in the history of the Solar System⁸.

The mineralogy of interstellar dust can be studied⁹⁻¹¹ by infrared spectroscopy, using the Infrared Space Observatory (ISO)¹². Here we concentrate on two evolved stars (NGC6302 and NGC6537, see Fig. 1): each of these stars has lost its envelope by means of a dusty stellar wind, usually referred to as a planetary nebula, and is about to fade away as a white dwarf. Both stars are known to have a circumstellar torus, containing the bulk of the dust mass^{13,14}.

We identify a strong emission band at a wavelength of 92.6 μm as being due to calcite (CaCO₃; Figs 1, 3), which is the only dust species known to us which has a feature at this wavelength. Dolomite (CaMg(CO₃)₂) and ankerite (CaFe(CO₃)₂) both have a band near 62 μm , which fits well with the observed spectrum of NGC6302 (Fig. 3), but unfortunately blends with the crystalline H₂O band at 62–63 μm and the diopside (CaMgSi₂O₆) band at 65 μm . The unambiguous identification of those two members of the dolomite group is therefore less secure. However, using dolomite as a representative for the group substantially improves the fit to the spectrum of NGC6302, and results in a reasonable abundance of crystalline H₂O ice (see below). Other carbonates containing cosmically abundant elements may also be present. However, magnesite (MgCO₃) and siderite (FeCO₃) do not have strong far-infrared bands, so their presence in the cold dust cannot be established. Aragonite (CaCO₃, orthorhombic) may be present, but it is formed at high pressure and temperature and therefore is not likely to occur in this environment. From Fig. 2 it is evident that

both calcite and dolomite also show strong bands at wavelengths below 92.6 μm that are comparable in strength with the 92.6- μm band. However, our model calculations show that the low temperature of the bulk of the dust (30–60 K) suppresses these shorter-wavelength bands. Moreover, the combined mass absorption coefficient at these wavelengths is largely determined by other more abundant dust species, making the carbonate contribution insignificant.

All solid-state features appear in emission at mid- and far-infrared wavelengths, so the dust must be optically thin at these wavelengths. The exact masses of the different components depend on the temperature of the dust. Assuming that all dust species have the same temperature distribution, we can estimate the mass ratio between the different dust components using the opacities measured in the laboratory. The Ca-bearing minerals—dolomite, calcite and diopside—together take up less than 1% of the cold dust mass while still being prominent in the spectrum. The warm component shows prominent carbon-rich dust features and lacks small silicate grains. It thus appears to have a composition which is different from that of the cold dust. Owing to the lack of the 6.8- and 11.4- μm features, we can rule out a substantial warm ($T = 100$ –118 K) carbonate component. Assuming solar abundances, we find that $\sim 30\%$ of the Ca is locked up in the modelled Ca-bearing minerals. The rest of the Ca is probably taken up by other solids that cannot be identified from the spectrum, such as amorphous silicates. We find that, assuming an H₂O/H₂ abundance ratio of 1×10^{-3} by mass¹⁵, only $\sim 10\%$ of the H₂O is in the form of ice. If dolomite and diopside are not included in the model, all water needs to be contained in water ice, which is hard to reconcile with observations of OH masers¹⁶ that require the presence of a significant amount of gas-phase water. Therefore, dolomite and diopside must contribute significantly to the 60- μm complex.

By analogy with terrestrial processes, the presence of carbonates in meteorites is interpreted as evidence for secondary processing on large parent bodies with liquid water, although an alternative origin has been suggested¹⁷. However, our detection of carbonates in planetary nebulae shows that the formation of carbonates may

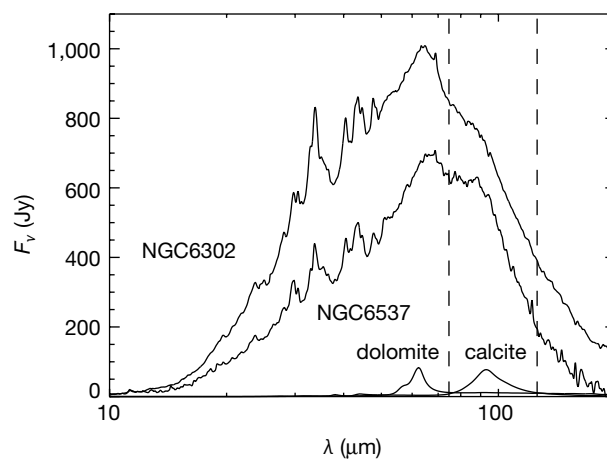


Figure 1 Comparison between the spectra of NGC6302 and NGC6537. The spectrum of NGC6537 is multiplied by a factor of 3.5. The modelled contributions of calcite and dolomite to NGC6302 are plotted. The dashed lines indicate the width of the calcite band for comparison. The spectra are dominated by a smooth continuum caused by amorphous silicates, and contain a wealth of emission bands²² due to crystalline silicates (forsterite (Mg₂SiO₄), enstatite (MgSiO₃) and diopside (CaMgSi₂O₆)^{23,24}) and crystalline H₂O ice^{25,26}. We use published laboratory data for these materials, as well as performing laboratory measurements of carbonates that cover the entire ISO wavelength region between 2 and 200 μm (Fig. 2) to fit the spectrum of NGC6302 (Fig. 3). λ , wavelength; F_ν , observed specific flux.

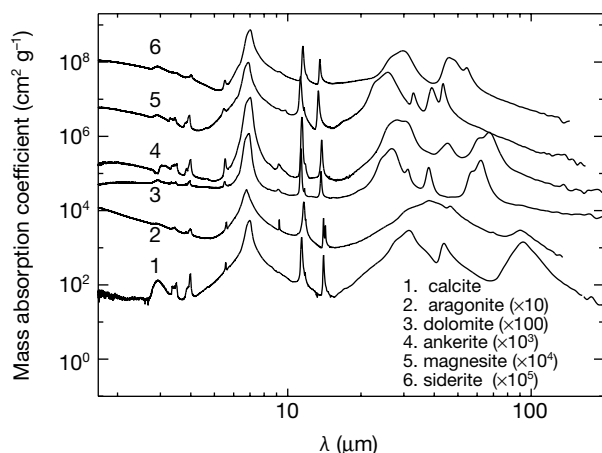


Figure 2 Mass absorption coefficients of different carbonates. These coefficients were derived from the transmission spectra of the samples. For transmission measurements, powders produced by grinding and sedimentation of the minerals were embedded in a transparent material (KBr and polyethylene). Spectroscopy was performed by means of a Bruker 113v Fourier transform infrared spectrometer in the wavenumber range from 5,000 to 50 cm^{-1} (2–200 μm). The spectral absorption features in the carbonate spectra arise mainly from the anionic carbonate group. The band at about 7 μm can be attributed to the asymmetric stretching vibration of C–O. The strong features at about 11 and 14 μm can be assigned to the out-of-plane and in-plane bending mode of the $(\text{CO}_3)^{2-}$ ion, respectively. The band at 92 μm detected in calcite is caused by a planar deformation of the carbonate (E_u) mode.

occur under much more primitive conditions. The age of the dust shell in the planetary nebula is $\sim 10^4$ years, which makes the process of parent-body accretion required for aqueous alteration, and the subsequent shattering into sub-micrometre sized grains, very unlikely. The mass of the ionized gas in the nebula is¹⁸ two solar masses, providing a lower limit to the total gas mass, which also includes neutral and molecular components. Assuming a gas-to-dust ratio of 100 and using the relative dust masses, we find that the mass of the carbonates is ≥ 30 Earth masses. This is far too large to be a relic of a hypothetical planetary system that may have formed together with the star that is now the central star of the planetary nebula, considering that abundance constraints imply that calcite and dolomite cannot make up more than $\sim 2\%$ of the total mass of rocky material. This leaves no other conclusion than that an alternative formation mechanism exists for the carbonates.

We discuss three alternative mechanisms. First, the coexistence of carbonates with water ice suggests that the water ice layer frozen out on the grain surface has an important role. It is known that a few layers of water molecules which are directly adjacent to the surface of the grain are quite mobile, with properties similar to liquid water¹⁹. If CO_2 molecules are included in the ice layer, a grain-surface reaction with Ca^{2+} or Mg^{2+} contained in a silicate lattice could possibly lead to the formation of carbonates. However, we do not find spectral evidence for a mixture of $\text{H}_2\text{O}/\text{CO}_2$ ice. Moreover, the thickness of the layer of mobile water molecules in the ice layer depends on the temperature, and is virtually absent below 193 K (ref. 19). The CO_2 would freeze out at a much lower temperature than water ice, causing a layered ice structure which would prevent interaction between the cations in the grain and the CO_2 , especially when the mobility is low.

Second, we propose a mechanism for carbonate formation through the reaction of H_2O and CO_2 in the gas phase with the grain surface. This mechanism proceeds through an intermediate step of the formation of hydrated silicates, which increases the availability of cations required for the synthesis of carbonates²⁰. This process occurs at a pressure of 10^{-4} atm and $T = 350$ K. However, in

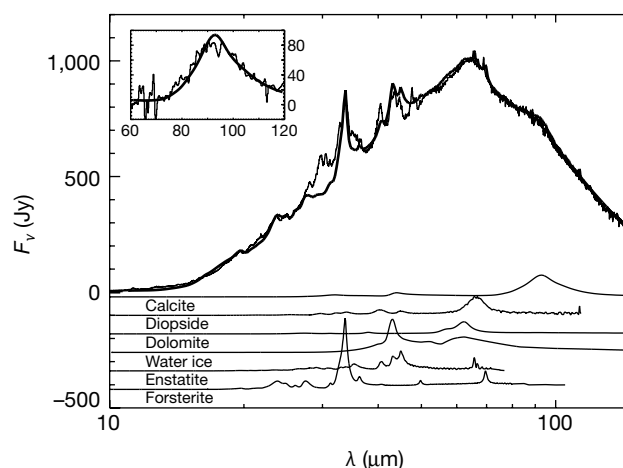


Figure 3 Model fit to the observed spectrum of planetary nebula NGC6302. The modelled spectrum is represented by the thick line, and the observations by the thin line. The inset shows the observed 92.6- μm feature (thin line), derived by subtracting all contributions of previously identified dust species from the observed spectrum, overplotted by the modelled contribution of calcite (thick line). The spectrum of NGC6302 is reproduced assuming that the dust shell is optically thin at the mid- and far-infrared wavelengths, and that all dust species are found in the same temperature range. We assumed a grain size of 0.1 μm , typical of stellar outflows. The dust components required to reproduce the observed spectrum are amorphous olivine²⁷, metallic iron²⁸, forsterite²³, clino-enstatite²³, water ice^{25,26}, diopside²⁴, calcite and dolomite. All contributions, except those of the featureless components amorphous olivine and metallic iron (or amorphous carbon), are included in the bottom part of the figure.

the outflow of evolved stars, the available pressures at these temperatures are ~ 7 – 8 orders of magnitude lower than this, making this mechanism unlikely. Moreover, there are no spectral indications for the presence of hydrated silicates.

Last, carbonates are stable below $\sim 1,200$ K, so they could already be present at the silicate condensation radius. As terrestrial sediments at impact sites suggest, calcite and dolomite can condense directly from a gas containing high abundances of CO_2 and MgO/CaO at $\sim 1,000$ – $1,200$ K (ref. 21). This process has not yet been reproduced in a laboratory experiment, but deserves a more careful examination because of its observed occurrence.

The presence of carbonates in the dust shells of planetary nebulae require a non-aqueous formation mechanism. This mechanism probably also played a role in the solar nebula, where physical conditions resembled those in the outflows of evolved stars. If carbonates are not exclusively formed in the presence of liquid water, as was previously thought, then carbonates found in interplanetary dust particles and chondritic meteorites do not unambiguously indicate aqueous alteration of the grains. □

Received 16 March; accepted 5 November 2001.

- Mittlefehldt, D. W. ALH84001, a cumulative orthopyroxenite member of the martian meteorite clan. *Meteoritics* **29**, 214–221 (1994).
- Scott, E. R. D., Krot, A. N. & Yamaguchi, A. Carbonates in fractures of Martian meteorite Allan Hills 84001: petrologic evidence for impact origin. *Meteorit. Planet. Sci.* **33**, 709–719 (1998).
- Penman, J. M. Measurements of infrared reflectivity of astronomically interesting non-silicates. *Mon. Not. R. Astron. Soc.* **176**, 539–545 (1976).
- Gillet, F. C., Forrest, W. J. & Merrill, K. M. 8–13 micron spectra of NGC7027, BD+30°3639, and NGC6572. *Astrophys. J.* **183**, 87–93 (1973).
- Sandford, S. A. & Walker, R. M. Laboratory infrared transmission spectra of individual interplanetary dust particles from 2.5 to 25 microns. *Astrophys. J.* **291**, 838–851 (1985).
- Keane, J. V., Tielens, A. G. G. M., Boogert, A. C. A., Schutte, W. A. & Whittet, D. C. B. Ice absorption features in the 5–8 μm region toward embedded protostars. *Astron. Astrophys.* **376**, 254–270 (2001).
- Cohen, M. *et al.* The infrared emission bands. I. Correlation studies and the dependence on C/O ratio. *Astrophys. J.* **302**, 737–749 (1986).
- Endress, M., Zinner, E. & Bischoff, A. Early aqueous activity on primitive meteorite parent bodies. *Nature* **379**, 701–703 (1996).
- Waters, L. B. F. M. *et al.* Mineralogy of oxygen-rich dust shells. *Astron. Astrophys.* **315**, L361–L364 (1996).

10. Jäger, C. *et al.* Steps toward interstellar silicate mineralogy IV. The crystalline revolution. *Astron. Astrophys.* **339**, 904–916 (1998).
11. Molster, F. J. *et al.* Low-temperature crystallization of silicate dust in circumstellar disks. *Nature* **401**, 563–565 (1999).
12. Kessler, M. F. *et al.* The Infrared Space Observatory (ISO) mission. *Astron. Astrophys.* **315**, L27–L31 (1996).
13. Lester, D. F. & Dinerstein, H. L. An infrared disk at the center of the bipolar planetary nebula NGC 6302. *Astron. Astrophys. J.* **281**, L67–L69 (1984).
14. Corradi, R. L. M. & Schwarz, H. E. The kinematics of the high velocity bipolar nebulae NGC 6537 and HB 5. *Astron. Astrophys.* **269**, 462–468 (1993).
15. Gonzalez-Alfonso, E. & Cernicharo, J. The water vapor abundance in circumstellar envelopes. *Astrophys. J.* **525**, 845–862 (1999).
16. Payne, H. E., Phillips, J. A. & Terzian, Y. A young planetary nebula with OH molecules—NGC 6302. *Astrophys. J.* **326**, 368–375 (1988).
17. Metzler, K., Bisschoff, A. & Stoeffler, D. Accretionary dust mantles in CM chondrites—Evidence for solar nebula processes. *Geochim. Cosmochim. Acta* **56**, 2873–2897 (1992).
18. Pottasch, S. R. & Beintema, D. A. The ISO spectrum of the planetary nebula NGC 6302. II. Nebular abundances. *Astron. Astrophys.* **347**, 975–982 (1999).
19. Rietmeijer, F. J. M. A model for diagenesis in proto-planetary bodies. *Nature* **313**, 293–294 (1985).
20. Lancet, M. S. & Anders, E. Carbon isotope fractionation in the Fischer-Tropsch synthesis and in meteorites. *Science* **170**, 980–982 (1970).
21. Pope, K. O., Ocampo, A. C., Fischer, A. G., Morrison, J. & Sharp, Z. Carbonate condensates in the Chicxulub ejecta deposits from Belize. *Lunar Planet. Sci.* **27**, 1045–1046 (1996).
22. Molster, F. J. *et al.* The complete ISO spectrum of NGC6302. *Astron. Astrophys.* **372**, 165–172 (2001).
23. Koike, C. *et al.* The spectra of crystalline silicates in infrared region. *Proc. 32nd ISAS Lunar Planet. Symp.* **32**, 175–178 (1999).
24. Koike, C. *et al.* Absorption spectra of Mg-rich Mg-Fe and Ca pyroxenes in the mid- and far-infrared regions. *Astron. Astrophys.* **363**, 1115–1122 (2000).
25. Bertie, J. E., Labbé, H. J. & Whalley, E. Absorptivity of Ice I in the range 4000–30 cm⁻¹. *J. Chem. Phys.* **50**, 4501–4520 (1969).
26. Warren, S. G. Optical constants of ice from the ultraviolet to the microwave. *Appl. Opt.* **23**, 1206–1225 (1984).
27. Jäger, C., Mutschke, H., Begemann, B., Dorschner, J. & Henning, Th. Steps toward interstellar silicate mineralogy I. Laboratory results of a silicate glass of mean cosmic composition. *Astron. Astrophys.* **292**, 641–655 (1994).
28. Henning, Th. & Stognienko, R. Dust opacities for protoplanetary accretion disks: influence of dust aggregates. *Astron. Astrophys.* **311**, 291–303 (1996).

Acknowledgements

This work is based on observations with ISO, an ESA project with instruments funded by ESA member states (especially the PI countries: France, Germany, the Netherlands and the UK) and with the participation of ISAS and NASA. We thank F. J. M. Rietmeijer, D. Fabian, J. Bouwman, C. Dominik, A. G. G. M. Tielens, J. Bradley, W. Schutte, J. Keane, P. Morris, L. P. Keller, H. Y. McSweeney Jr and R. N. Clayton for support and discussions. We acknowledge support from an NWO 'Pionier' grant.

Correspondence and requests for materials should be addressed to E.K. (e-mail: ciska@science.uva.nl).

Quantum states of neutrons in the Earth's gravitational field

Valery V. Nesvizhevsky*, Hans G. Börner*, Alexander K. Petukhov*, Hartmut Abele†, Stefan Baessler†, Frank J. Rueß†, Thilo Stöferle‡, Alexander Westphal‡, Alexei M. Gagarski‡, Guennady A. Petrov‡ & Alexander V. Strelkov§

* Institute Laue-Langevin, 6 rue Jules Horowitz, Grenoble F-38042, France

† University of Heidelberg, 12 Philosophenweg, Heidelberg D-69120, Germany

‡ Petersburg Nuclear Physics Institute, Orlova Roscha, Gatchina, Leningrad reg. R-188350, Russia

§ Joint Institute for Nuclear Research, Dubna, Moscow reg. R-141980, Russia

The discrete quantum properties of matter are manifest in a variety of phenomena. Any particle that is trapped in a sufficiently deep and wide potential well is settled in quantum bound states. For example, the existence of quantum states of electrons in an electromagnetic field is responsible for the structure of atoms¹⁶, and quantum states of nucleons in a strong nuclear field give rise to the structure of atomic nuclei¹⁷. In an analogous way, the gravitational field should lead to the formation of quantum states. But the gravitational force is extremely weak compared to the

electromagnetic and nuclear force, so the observation of quantum states of matter in a gravitational field is extremely challenging. Because of their charge neutrality and long lifetime, neutrons are promising candidates with which to observe such an effect. Here we report experimental evidence for gravitational quantum bound states of neutrons. The particles are allowed to fall towards a horizontal mirror which, together with the Earth's gravitational field, provides the necessary confining potential well. Under such conditions, the falling neutrons do not move continuously along the vertical direction, but rather jump from one height to another, as predicted by quantum theory^{1–3}.

In order to allow for the experimental observation of gravitational bound states, all interactions of the matter particles with other fields must be so small that their interference with the gravitational quantum phenomena can be neglected. The choice of neutrons seemed to us most favourable because (1) they are neutral, (2) they have a long lifetime, and (3) they are elementary particles with low mass. The reasons why these properties are advantageous will become more evident from the following explanations.

We now consider how to demonstrate that bound states exist for neutrons trapped in the Earth's gravitational field. The gravitational field alone does not create a potential well, as it can only confine particles by forcing them to fall along gravity field lines. We need a second 'wall' to create the well. This can be obtained by introducing a reflecting mirror. Let us consider a neutron, which is lifted up by a few millimetres and is then dropped vertically onto the mirror. The neutron wave is reflected by the mirror and interferes with itself, as illustrated in Fig. 1. This self-interference creates a standing wave in the neutron density: the probability of finding a neutron at a given height exhibits maxima and minima along the vertical direction, the position of which depends on the quantum number of the bound states. The neutron that was dropped has gone through quantum 'steps'.

The classical analogue, the vibrating musical string, gives a visualization of a particle in a rectangular potential well. In this case, strict boundary conditions must be met: both the wavefunction amplitudes of the particles and the displacement amplitudes of the string vanish at the boundaries. In contrast, the gravitational well described above is asymmetric: whereas the reflecting mirror (under total reflection condition) corresponds to an infinite sharp

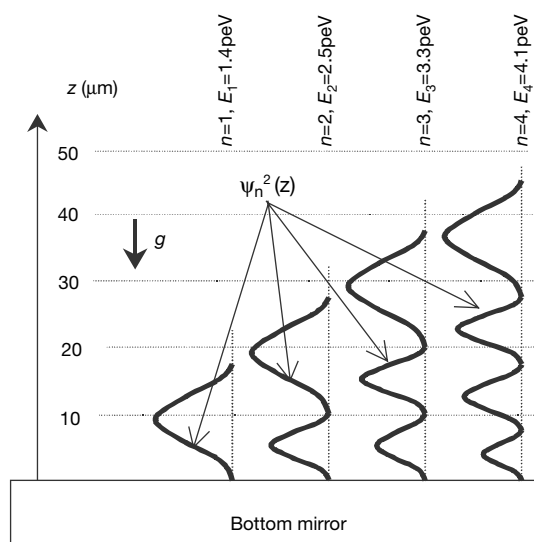


Figure 1 Wavefunctions of the quantum states of neutrons in the potential well formed by the Earth's gravitational field and the horizontal mirror. The probability of finding neutrons at height z , corresponding to the n th quantum state, is proportional to the square of the neutron wavefunction $\psi_n^2(z)$. The vertical axis z provides the length scale for this phenomenon. E_n is the energy of the n th quantum state.

wall, the gravitational field is much softer; as a result, the gravitational well extends in the opposite direction to the gravity field with increasing quantum number. Consequently, as can be seen in Fig. 1, the neutron wavefunctions are deformed upwards, and the energy differences between states with neighbouring quantum numbers become smaller with increasing level number. More detailed discussions, precise analytical solutions and related publications can be found elsewhere^{1–10}. We will here simply summarize the results of these analyses: the energies of the four lowest quantum states of a neutron in the Earth's gravitational field are $E_1 \approx 1.41$ peV, $E_2 \approx 2.46$ peV, $E_3 \approx 3.32$ peV and $E_4 \approx 4.08$ peV, respectively ($1 \text{ peV} = 10^{-12} \text{ eV}$). It is worth keeping in mind that the classical energy that is needed to lift a neutron by $10 \mu\text{m}$ against gravitation on Earth (given by mgz) is almost exactly 1 peV. (Here m is the neutron mass, g is the acceleration due to gravity, and z is the height.) Thus the energy E_1 corresponds in the classical approximation to the height $z_1 \approx 15 \mu\text{m}$, at which the first level of the quantum phenomenon for neutrons should be observed. This 'macroscopic' height is very advantageous, and helps us to design experiments to demonstrate the existence of gravitational levels for neutrons.

In a realistic experiment it is not possible to just lift a neutron, let it drop, and then measure its density distribution as a function of height. But we can take a beam of neutrons and let them fly horizontally above a reflecting mirror. If all forces can be eliminated except for gravitation and repulsion by the mirror (such as that due to magnetic field gradients, mechanical vibrations and so on), then the motion of the neutrons can be decoupled into independent vertical and horizontal components. The gravitational force acts on the vertical component only, and in this direction we then obtain the potential well that leads to the consequences described above. No forces act on the horizontal velocity component.

In order to further characterize our experiment, we have to make use of the uncertainty principle, which relates the Planck constant ($\hbar = 6.6 \times 10^{-16} \text{ eVs}$) to the minimal time period $\Delta\tau$, during which quantum states can be resolved with an energy difference ΔE : $\Delta\tau \approx \hbar/\Delta E$. Therefore, the vertical energy scale of the quantum levels $E_1 \approx 1.41$ peV requires that $\Delta\tau \gg 0.5 \text{ ms}$. A compromise has to be found between the length of the reflecting mirror and the horizontal velocity of neutrons. We have used mirrors with a length of 10 cm , and neutrons with a velocity of $\sim 10 \text{ m s}^{-1}$. A powerful source of ultracold neutrons (UCN)^{11–14}, which operates at the Institute Laue-Langevin, Grenoble¹⁵, provides the neutrons with such velocity. The energy E_1 corresponds in the classical approximation to the vertical velocity component $v \approx 1.7 \text{ cm s}^{-1}$, which is significantly smaller than the horizontal velocity component. If we let the neutrons fly 'slightly upwards' (see Fig. 2), they will follow a parabolic trajectory due to gravity. At the maximum of the parabola their vertical velocity component will be zero in the classical approximation, and will then increase again. To limit the vertical velocity component, we use an absorber parallel to the bottom mirror and placed above it (see Fig. 2). The distance between absorber and mirror can be adjusted.

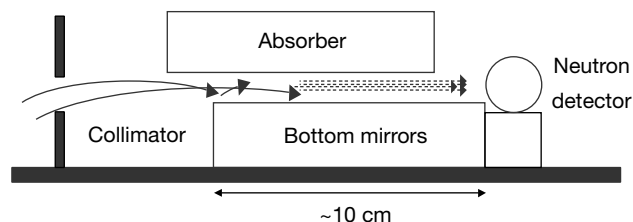


Figure 2 Layout of the experiment. The limitation of the vertical velocity component depends on the relative position of the absorber and mirror. To limit the horizontal velocity component we use an additional entry collimator. The relative height and size of the entry collimator can be adjusted.

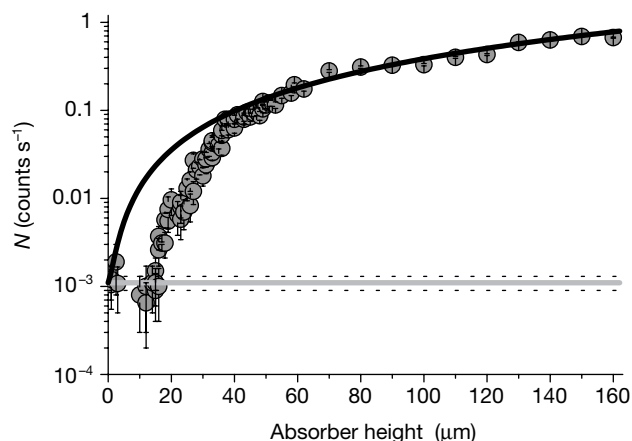


Figure 3 The neutron throughput N versus the absorber height. The circles represent the data points. The solid curve is the classical fit to these data. The slit becomes transparent to neutrons at a finite slit opening. The horizontal straight lines indicate the detector background values and uncertainties measured while the neutron source was 'off'.

In our experiments, neutrons flow between the mirror below and absorber above, and the neutron transmission N is measured as a function of the width Δz of the slit defined by the mirror and the absorber. This width Δz acts as a selector for the vertical velocity component. In order to keep the vertical and horizontal velocity components decoupled, severe restrictions concerning quality and adjustment of the different parts used in the set-up must be met^{2,3}. Ideally, from Fig. 1, we expect a stepwise dependence of N as a function of Δz . If Δz is smaller than the spatial width of the lowest quantum state, then N should be zero. When Δz is equal to the spatial width of the lowest quantum state, then N should increase sharply. Further increase in Δz should not increase N as long as Δz is smaller than the spatial width of the second quantum state. Then, N should again increase stepwise. At sufficiently large slit width Δz , the classical dependence $N \sim \Delta z^{1.5}$ should be approached, and the stepwise increase should be washed out. (Naively we might expect that classically $N \sim \Delta z$; this is not the case because we obtain an additional $z^{0.5}$ due to the fact that an increase in Δz also allows for an increase in the accepted spread of velocities.) The identification of the lowest quantum state is easier than that of the higher states because in this case the relative change in the count rate N is maximal.

The effects that we observe, shown in Fig. 3, are consistent with the expectations described above. In particular, the non-transparency of the slit (formed by the bottom mirror and the absorber) is clearly observed for the neutrons when the slit width is smaller than the spatial width of the lowest quantum state. We note that the 'diameter' of a neutron is $\sim 10^{-13} \text{ cm}$, which is much smaller than the width of $\geq 15 \mu\text{m}$ at which the slit starts to become transparent for neutrons. Careful analysis of the experiment has allowed us to rule out any systematic errors. In particular, tests have shown that the shape of the transmission curve (Fig. 3) does not depend on the value of the horizontal velocity component, but that it depends only on the vertical velocity component, as expected. If the slit is opened up to $15 \mu\text{m}$, it is just not transparent for neutrons. But it is sufficiently large that we can observe transmission of visible light, although the wavelength of light ($\sim 0.6 \mu\text{m}$) is much larger than the neutron wavelength of $\sim 0.01 \mu\text{m}$; this observation tells us that the slit is really open and well adjusted. Evidently, the difference in transmission results from the fact that the Earth's gravitational field does not act noticeably on photons within the frame of our experimental set-up.

Figure 4 shows on an extended scale the initial part of the transmission curve N as a function of slit width Δz . The dashed

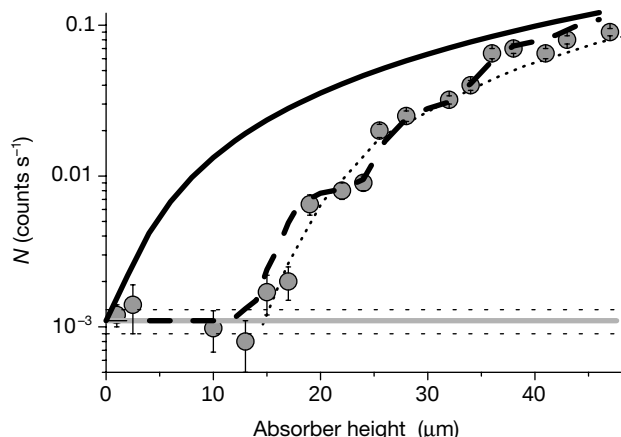


Figure 4 The neutron throughput versus the absorber height at low height values. The data points are summed up in intervals of 2 μm . The dashed curve corresponds to a fit using the quantum-mechanical calculation, in which all level populations and the height resolution are fitted from the experimental data. The solid curve is again the full classical treatment. The dotted line is a truncated fit in which it is assumed that only the lowest quantum state—which leads to the first step—exists.

curve shows the results of a quantum fit, in which the level populations and the height resolution are free parameters. The solid line is again the full classical treatment ($N \sim z^{1.5}$). The dotted line is a truncated fit to the assumption that only the lowest quantum level—which leads to the first step—exists. Then it continues at the absorber height of $z_1 \approx 15 \mu\text{m}$ with a shifted classical treatment ($N \sim (z - z_1)^{1.5}$) that is more like a ‘guide to the eye’ curve. Our statistics for large slit width are still not sufficient, but the existence of the first step due to the lowest quantum level is clearly reproduced.

Our experimental observations of the neutron quantum states in the Earth’s gravitational field provide another demonstration of the universality of the quantum properties of matter, but at this stage we have only shown a phenomenon that was expected—although not easy to prove. As the parameters of quantum states are defined in such a system mainly by the interaction of the neutron with the gravitational field, the phenomenon we report can now be considered for further investigations of fundamental properties of matter. Thus, as it is evident from the uncertainty principle, the energy resolution ΔE could be improved significantly by increasing $\Delta \tau$ (in principle, ΔE could be as low as $\sim 10^{-18}$ eV if $\Delta \tau$ approaches the lifetime of the neutron, so that the level width becomes a million times smaller than the energy difference between levels). The use of resonance transitions between such narrow levels could find applications in physics, such as the precise verification of the proportionality of inertial and gravitational masses of elementary particles (neutrons), and a check of the electrical neutrality of neutrons—which is not a trivial fact. Increasing the time that neutrons spend in the gravitational bound states will become one of the main challenges in extending this experiment. When trying to achieve this, it will be necessary to demonstrate that the neutrons are spending a much longer time in the potential well, and a significant increase in the available density of ultracold neutrons will be necessary. □

Received 10 October; accepted 22 November 2001.

1. Lushchikov, V. I. & Frank, A. I. Quantum effects occurring when ultracold neutrons are stored on a plane. *JETP Lett.* **28**, 559–561 (1978).
2. Nesvizhevsky, V. V. *et al.* Search for quantum states of the neutron in a gravitational field: gravitational levels. *Nucl. Instrum. Methods Phys. Res.* **440**, 754–759 (2000).
3. Nesvizhevsky, V. V. *et al.* in *ILL Annual Report* (eds Cicognani, G. & Vettier, Ch.) 64–65 (Institute Laue-Langevin, Grenoble, 2000).
4. Landau, L. D. & Lifshitz, E. M. *Quantum Mechanics* 164–196 (Pergamon, Oxford, 1976).
5. Flügge, S. *Practical Quantum Mechanics* (Mir, Moscow, 1974).
6. Colella, R. A., Overhauser, W. & Werner, W. A. Observation of gravitationally induced quantum interference. *Phys. Rev. Lett.* **34**, 1472–1474 (1975).

7. Baryshevskii, V. G., Cherepizta, S. V. & Frank, A. I. Neutron spin interferometry. *Phys. Lett. A* **153**, 299–302 (1991).
8. Frank, A. I. Modern optics of long-wavelength neutrons. *Sov. Phys. Usp.* **34**, 980–987 (1991).
9. Felber, J., Gähler, R., Rauch, C. & Golub, R. Matter waves at a vibrating surface: Transition from quantum-mechanical to classical behavior. *Phys. Rev. A* **53**, 319–328 (1996).
10. Peters, A., Chung, K. Y. & Chu, S. Measurement of gravitational acceleration by dropping atoms. *Nature* **400**, 849–852 (1999).
11. Lushchikov, V. I., Pokotilovskiy, Yu. N., Strelkov, A. V. & Shapiro, F. I. Observation of ultracold neutrons. *JETP Lett.* **9**, 23–26 (1969).
12. Steyerl, A. Measurement of total cross sections for very slow neutrons with velocities from 100 m/s to 5 m/s. *Phys. Lett. B* **29**, 33–35 (1969).
13. Ignatovich, V. K. *The Physics of Ultracold Neutrons* (Clarendon, Oxford, 1990).
14. Golub, R., Richardson, D. J. & Lamoreux, S. K. *Ultracold Neutrons* (Higler, Bristol, 1991).
15. Steyerl, A. & Malik, S. S. Sources of ultracold neutrons. *Nucl. Instrum. Methods Phys. Res. A* **284**, 200–207 (1989).
16. Born, M. *Atomic Physics* (Blackie & Son, London, 1969).
17. Bohr, A. & Mottelson, B. R. *Nuclear Structure* (Benjamin, New York, 1969).

Acknowledgements

We are grateful to our colleagues who were interested in this research and contributed to its development, in particular K. Ben-Saidane, D. Berruyer, Th. Brenner, J. Butterworth, D. Dubbers, P. Geltenbort, T. M. Kuzmina, A. J. Leadbetter, B. G. Peskov, S. V. Pinaev, K. Protasov, I. A. Snigireva, S. M. Soloviev and A. Voronin. The work was supported by INTAS.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to V.V.N.
(e-mail: nesvizhevsky@ill.fr).

Antiferromagnetic order induced by an applied magnetic field in a high-temperature superconductor

B. Lake^{*,†}, **H. M. Rønnow**[‡], **N. B. Christensen**^{§,||}, **G. Aeppli**^{§,||}, **K. Lefmann**[§], **D. F. McMorrow**[§], **P. Vorderwisch**[‡], **P. Smeibidl**[‡], **N. Mangkorntong**[☆], **T. Sasagawa**[☆], **M. Nohara**[☆], **H. Takagi**[☆] & **T. E. Mason**^{**}

^{*} Oak Ridge National Laboratory, PO Box 2008 MS 6430, Oak Ridge, Tennessee 37831-6430, USA

[†] Department of Condensed Matter Physics, University of Oxford, Clarendon Laboratory Parks Road, Oxford OX1 3PU, UK

[‡] CEA (MDN/SPSMS/DRFMC), 17 Ave. des Martyrs, 38054 Grenoble cedex 9, France

[§] Materials Research Department, Risø National Laboratory, 4000 Roskilde, Denmark

^{||} Ørsted Laboratory, Niels Bohr Institute for APG, Universitetsparken 5, DK 2100, Copenhagen, Denmark

[¶] NEC Research Institute, 4 Independence Way, Princeton, New Jersey 08540-6634, USA

[#] BENSCH, Hahn-Meitner Institut, Glienicke Strasse 100, 14109 Berlin, Germany

[☆] Department of Advanced Materials Science, Graduate School of Frontier Sciences, University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-8656, Japan

^{**} Experimental Facilities Division, Spallation Neutron Source, 701 Scarborough Road, Oak Ridge, Tennessee 37830, USA

One view of the high-transition-temperature (high- T_c) copper oxide superconductors is that they are conventional superconductors where the pairing occurs between weakly interacting quasiparticles (corresponding to the electrons in ordinary metals), although the theory has to be pushed to its limit¹. An alternative view is that the electrons organize into collective textures (for example, charge and spin stripes) which cannot be ‘mapped’ onto the electrons in ordinary metals. Understanding the properties of the material would then need quantum field theories of objects

such as textures and strings, rather than point-like electrons^{2–6}. In an external magnetic field, magnetic flux penetrates type II superconductors via vortices, each carrying one flux quantum⁷. The vortices form lattices of resistive material embedded in the non-resistive superconductor, and can reveal the nature of the ground state—for example, a conventional metal or an ordered, striped phase—which would have appeared had superconductivity not intervened, and which provides the best starting point for a pairing theory. Here we report that for one high- T_c superconductor, the applied field that imposes the vortex lattice also induces ‘striped’ antiferromagnetic order. Ordinary quasiparticle models can account for neither the strength of the order nor the nearly field-independent antiferromagnetic transition temperature observed in our measurements.

$\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ is the simplest high- T_c superconductor. The undoped compound is an insulating antiferromagnet, where the spin moments on adjacent Cu^{2+} ions are antiparallel⁸. Introduction of charge carriers via Sr doping reduces the ordered moment until it vanishes at $x < 0.13$. In addition, for $x > 0.05$ the commensurate antiferromagnetism is replaced by incommensurate order^{2,3,9,10}, where the repeat distance for the pattern of ordered moments is substantially larger than the spacing between neighbouring copper ions. $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ becomes a superconductor for Sr dopings of $0.06 < x < 0.25$, and for underdoped compounds ($0.06 < x < 0.13$), both superconductivity and magnetic order are present. Figure 1a shows the electrical resistivity as a function of temperature T and magnetic field H , applied perpendicularly to the superconducting

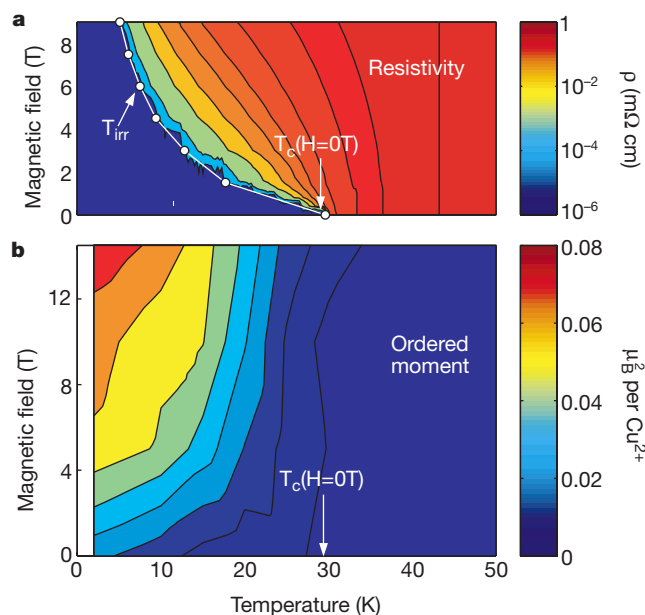


Figure 1 Magneto-transport and neutron diffraction data for $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ as a function of temperature and magnetic field. **a**, Magneto-transport measurements parallel to the CuO_2 planes, obtained via a standard four-probe method; the colours indicate the electrical resistivity ρ . In a magnetic field, the sharp transition from normal to superconducting states is broadened into a crossover region and vortices are thought to form at temperatures where the resistivity falls below its value at $T_c(H=0\text{ T})$. Phase coherent superconductivity, characterized by zero-resistivity sets in at the much lower ‘irreversibility’ temperature ($T_{\text{irr}}(H)$), marked by the white circles. **b**, The square of the ordered spin moment per Cu^{2+} ion as a function of temperature and applied magnetic field. The ordered moment squared is proportional to the observed neutron-scattering signal, and was deduced from scans similar to those shown in Fig. 2. It first becomes significant below the zero-field superconducting transition temperature ($T_c(H=0\text{ T})$), and increases with decreasing temperature and increasing field. The crystals used for these measurements were grown in an optical image furnace.

CuO_2 planes, for a sample with $x = 0.10$. The resistivity decreases rapidly below the zero-field transition temperature $T_c(H=0\text{ T}) = 29\text{ K}$, and becomes zero below the irreversibility temperature $T_{\text{irr}}(H)$. The white circles locate T_{irr} and show that it is a rapid function of applied field, so that even for fields much smaller than the upper critical field ($H_{c2}(x=0.10) \approx 45\text{ T}$; ref. 11), perfect conductivity does not occur until the temperature is well below $T_c(H=0\text{ T})$.

In our experiments we used magnetic neutron diffraction to measure the spin ordering in single crystals of underdoped $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ ($x=0.10$). The technique is analogous to mapping the positions of atoms in crystals using X-ray or neutron diffraction. The superconducting CuO_2 planes were aligned in the scattering plane, and the magnetic field was applied perpendicular to these planes. The inset in Fig. 2a shows the reciprocal space region over which data were collected. In zero field (Fig. 2a), incommensurate elastic peaks occur in the superconducting phase at low-temperatures. The peaks are resolution-limited, implying an in-plane correlation length of $\zeta > 400\text{ Å}$, similar to that observed¹⁰ for

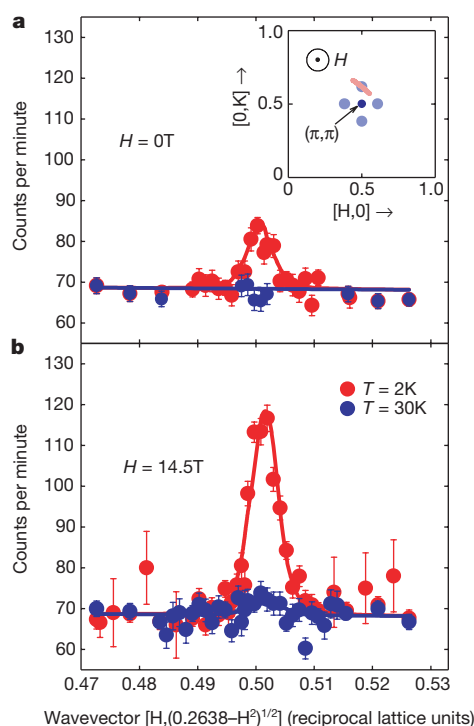


Figure 2 Magnetic neutron diffraction data for $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ with $x = 0.10$. The inset shows the relevant reciprocal space, labelled using the two-dimensional notation appropriate for the superconducting CuO_2 planes. The black dot at $(0.5, 0.5)$ represents the Bragg point associated with the commensurate antiferromagnetism of the insulating $x = 0$ parent compound. The incommensurate antiferromagnetic order in our metallic $x = 0.10$ material gives rise to diffraction at the quartet of blue dots¹⁰. The pink line shows the trajectory of a typical scan which passes through the incommensurate peak at $(0.5, 0.6175)$ and the direction of the applied field is shown to be perpendicular to the CuO_2 planes. **a**, The data collected in zero field. The red circles show that signal is present in the superconducting phase at $T = 1.9\text{ K}$. But just above the superconducting transition temperature at $T = 30\text{ K}$, the signal has completely disappeared (blue circles). **b**, Data measured at the same temperatures as **a**, but in a field of $H = 14.5\text{ T}$; the low-temperature signal is a factor of three larger than the zero-field signal. The data were collected using the V2/FLEX neutron scattering spectrometer at the BER II reactor, Hahn-Meitner Institute, Berlin. $60'$ collimators were used to define the divergence of the neutron beam. The energy was fixed at 7.5 meV using a pyrolytic graphite (PG) monochromator and analyser, and a tunable PG filter was used to eliminate second-order scattering.

$x = 0.12$. The peak amplitude decreases as T is increased, and is entirely absent at $T = 30$ K. Our new finding is that external magnetic fields markedly increase the low-temperature signal. Figure 2b shows that for $H = 14.5$ T and $T = 2$ K the signal is three times larger than in zero field, while at 30 K, just above $T_c(H = 0$ T), there is very little field-induced signal. The field-induced signal is resolution-limited, and the magnetic in-plane correlation length ($\xi > 400$ Å), is much greater than the superconducting coherence length ($\xi \approx 20$ Å) and the inter-vortex spacing ($a_v = 130$ Å for $H = 14.5$ T). As the superconducting coherence length gives the size of the vortices, coherent magnetism cannot reside in the vortex cores alone, but must extend across the superconducting regions of the material. This result is important because, taken together with the strong antiferromagnetic proximity effect which follows from the large zero-frequency antiferromagnetic susceptibility of superconducting $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ (ref. 12), it implies that as long as the vortex array is not radically different from a conventional Abrikosov lattice, superconductivity and antiferromagnetism coexist throughout the bulk of the material.

We collected scans similar to those shown in Fig. 2 for a variety of fields and temperatures. The neutron signal is normalized using a standard phonon-based calibration to yield the ordered spin moment squared, as displayed in Fig. 1b. Order develops just below the zero-field transition temperature $T_c(H = 0$ T), and increases with decreasing temperature and increasing field. Figure 3a shows the temperature dependence; the zero-field order (blue circles) increases gradually below $T_c(H = 0$ T), reaching a maximum of $0.15 \pm 0.01 \mu_B$ per Cu^{2+} ion (μ_B is the Bohr magneton and the error is statistical). This value is somewhat larger than the value of $0.10 \pm 0.04 \mu_B$ per Cu^{2+} deduced from neutron-diffraction

measurements for a sample with $x = 0.12$ (ref. 13). The field-induced order for $H = 5$ T, 10 T and 14.5 T is also plotted; it was obtained by subtracting the zero-field signal from the signal measured in-field. The temperature dependence of the field-induced signal is clearly different from that for $H = 0$ T. It increases more rapidly just below $T_c(H = 0$ T) and saturates for $T < 10$ K, where in zero field the moment is still evolving. Figure 3b shows the H dependence of the field-induced order at base temperature ($T = 1.9$ K). The order increases rapidly with small fields, probably reflecting the linear dependence of vortex density on field. The order increases more slowly for higher fields as the antiferromagnetic regions, nucleated by the vortices, start to merge.

Our work represents a departure from two previous experiments^{14,15} on $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ in magnetic fields. In the first experiment, the static magnetism in a sample with $x = 0.12$ was found to undergo an enhancement in an applied magnetic field consistent with our findings¹⁴. But this measurement was confined to a single field (10 T) and temperature (4.2 K), and the sample was atypical—with an anomalously low value of $T_c(H = 0$ T) = 12 K. The zero-field Néel temperature, $T_N = 25$ K, was well above $T_c(H = 0$ T), and the onset temperature for the field-induced enhancement was unknown. The second experiment focused on magnetic dynamics in optimally doped $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ ($x = 0.163$) (ref. 15). This material showed no static magnetism in either zero or non-zero fields up to 14.5 T. But an applied field had the effect of recovering some of the magnetic fluctuations below the ‘spin gap’ which are suppressed by superconductivity in zero field. An important feature of this experiment was that the antiferromagnetic correlations developed at a temperature below T_{irr} ; that is, they only occurred once phase coherent superconductivity had been attained. Thus the results of ref. 15 are in some ways the inverse or ‘dual’ of those we report here on our sample with $x = 0.1$, where the static antiferromagnetism arises first and phase coherent superconductivity is established at a lower temperature, when the ordered moments have saturated near their base temperature values.

We believe that the field-induced signal is intrinsic, and arises from fundamental physical processes, even though the zero-field signal may well be stabilized by the disorder inherent in a random alloy such as $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$. Thus, the magnetic field does not simply enhance the existing antiferromagnetism but induces magnetic order in its own right. There are several reasons to support our belief. First, the sample is of very high quality, as shown by specific heat measurements¹⁶ and our transport data (Fig. 1a). Second, a different and poorer-quality sample, also with $x = 0.10$, which displays static antiferromagnetism in zero field above as well as below $T_c(H = 0$ T), showed field-induced order with exactly the same temperature dependence as reported here¹⁷. Third, the degree of order is substantial, corresponding to a volume fraction of $\sim 50\%$ for the incommensurate magnetic state, assuming that the ordered moment is $0.4\mu_B$ per Cu^{2+} , as suggested by muon spin relaxation data¹⁸. Last, whereas the zero-field signal indeed displays the gradual rise typically associated with defect-induced magnetism in correlated Fermi systems¹⁹, the field-induced signal increases according to classical mean-field theory, suggesting a different and intrinsic mechanism (Fig. 3a).

Our data bear on two popular descriptions of copper oxide superconductors. The first states that the copper oxides are describable in the same terms as ordinary solids, where small variations of key parameters lead to Fermi-surface instabilities corresponding to superconductivity and insulating ‘striped’ phases. Within such an interpretation, stripe order is simply tuned by doping, and appears because of Fermi-surface nesting and lattice anomalies near $x = 1/8$. However, what we have found is that below $x = 1/8$, very modest (on the scale of both electron and phonon energies) fields induce antiferromagnetic order, with a field-independent onset temperature and a strongly field-dependent amplitude. These results cannot

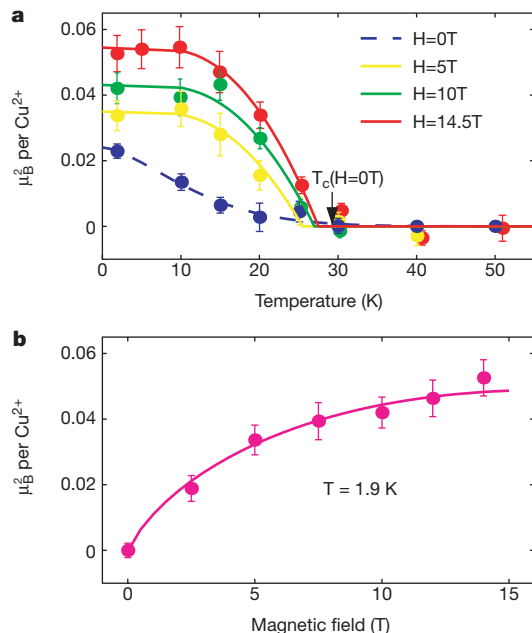


Figure 3 The dependence on temperature and field of the ordered spin moment squared. The data were calibrated using a transverse acoustic phonon measured around the (1, 1) Bragg peak (energy transfer $E = 2$ meV and sound velocity = 26.9 meV Å), and are presented in units of μ_B^2 per Cu^{2+} . **a**, The temperature dependence. The zero-field signal (blue circles) increases gradually below $T_c(H = 0$ T), and the dashed line is a guide to the eye. Also plotted is the field-induced signal (equal to the signal measured in-field minus the zero-field signal) for fields of $H = 5$ T (yellow circles), 10 T (green circles) and 14.5 T (red circles). The solid lines through the data are fits to mean field theory²⁶. **b**, The field-induced signal as a function of field at $T = 2$ K (magenta circles). The solid line is a fit to the data of the expression, $M^2(H/H_{c2})\ln(H_{c2}/H)$ (ref. 23), with fitted parameter $M^2 = 0.12 \mu_B^2$ per Cu^{2+} .

follow from a simple Fermi 'quasiparticle' description, which anticipates that the Néel temperature will rise significantly as the base-temperature ordered moment increases.

A second description of high- T_c superconductivity does not focus on microscopic detail (unlike the quasiparticle-based theories), but is instead based on the hypothesis that the copper oxides possess hidden 'SO(5)' symmetry where magnetic and superconducting order parameters can be 'rotated' into each other. Calculations based on this hypothesis predict antiferromagnetic vortices in a magnetic field, and have had some success in modelling the data^{20–23}. The original work was performed assuming that the symmetry is exact²⁰. In this model the antiferromagnetism is confined to the vortex cores which have size $\sim \xi = 20 \text{ \AA}$, and there are no regions of the compound where superconductivity and antiferromagnetism exist simultaneously. Our data suggest that the antiferromagnetic regions are much bigger than this ($\xi > 400 \text{ \AA}$), and extend well beyond the vortex cores into the superconducting areas. Alternatively, there could be interactions between the vortices that lead to coherent long-range antiferromagnetic order without the magnetism spreading beyond the cores; but then the corresponding Néel temperature would be a strong function of the intervortex spacing, controlled by H —whereas the onset temperature in our experiments is approximately field-independent.

More recent computations motivated by our experiments have relaxed SO(5) symmetry to allow coexistence of superconductivity and magnetism, with the antiferromagnetic correlations extending beyond the vortex cores into the bulk of the crystal^{22,23}. Although the underlying continuum approximations may be called into question at high fields, such a model²³ gives quantitative results in good agreement with our data for $T = 0$, as shown in Fig. 3b. The physical insight here is that the magnetism of the vortex state manifests a magnetic quantum critical point very close to the superconductor in the phase diagram²⁴.

We now consider how our results relate to the wider understanding of high- T_c superconductivity. First, they provide clear evidence for intrinsic antiferromagnetism coexisting with superconductivity in the same sample. Because the relative amplitudes of the two types of order can be continuously tuned by a magnetic field using a fixed sample (rather than by changing the doping), experiments to examine the trade-offs between magnetism and superconductivity should become much more detailed and reliable. Second, our data taken together with resistivity measurements²⁵ strongly suggest that in a magnetic field large enough to destroy superconductivity ($H > H_{c2}$), an underdoped copper oxide would become an incommensurate, antiferromagnetic insulator. This means that if we want to construct a theory of superconductivity at $T = 0$ in underdoped copper oxides, the most obvious starting point is this antiferromagnetic insulator, and not—as in conventional theories based on a Bardeen–Cooper–Schrieffer approach—a metallic Fermi liquid of weakly interacting quasiparticles. □

Received 15 February; accepted 20 November 2001.

1. Monthoux, P., Balatsky, A. V. & Pines, D. Weak-coupling theory of high-temperature superconductivity in the antiferromagnetically correlated copper oxides. *Phys. Rev. B* **46**, 14803–14817 (1992).
2. Zaanen, J. & Gunnarsson, O. Charged magnetic domain lines and the magnetism of high- T_c oxides. *Phys. Rev. B* **40**, 7391–7394 (1989).
3. Tranquada, J. M. *et al.* Evidence for stripe correlations of spins and holes in copper-oxide superconductors. *Nature* **375**, 561–563 (1995).
4. Kivelson, S. A., Fradkin, E. & Emery, V. J. Electronic liquid-crystal phases of a doped Mott insulator. *Nature* **393**, 550–553 (1998).
5. Zaanen, J. Stripes defeat the Fermi liquid. *Nature* **404**, 714–715 (2000).
6. Zaanen, J., Osman, O. Y. & van Saarloos, W. Metallic stripes: separation of spin, charge, and string fluctuation. *Phys. Rev. B* **58**, R11868–R11871 (1998).
7. Kittel, C. *Introduction to Solid State Physics* 6th edn, 317–358 (Wiley & Sons, New York, 1986).
8. Vaknin, D. *et al.* Antiferromagnetism in La_2CuO_4 . *Phys. Rev. Lett.* **58**, 2802–2805 (1987).
9. Lee, Y. S. *et al.* Neutron-scattering study of spin-density wave order in the superconducting state of excess-oxygen-doped $\text{La}_2\text{CuO}_{4+y}$. *Phys. Rev. B* **60**, 3643–3654 (1999).
10. Kimura, H. *et al.* Neutron-scattering study of static antiferromagnetic correlations in $\text{La}_{2-x}\text{Sr}_x\text{Cu}_{1-y}\text{Zn}_y\text{O}_4$. *Phys. Rev. B* **59**, 6517–6523 (1999).

11. Andoh, Y. *et al.* Resistive upper critical fields and irreversibility lines of optimally doped high- T_c cuprates. *Phys. Rev. B* **60**, 12475–12479 (1999).
12. Lake, B. *et al.* Spin gap and magnetic coherence in a clean high-temperature superconductor. *Nature* **400**, 43–46 (1999).
13. Wakimoto, S., Birgeneau, R. J., Lee, Y. S. & Shirane, G. Hole concentration dependence of the magnetic moment in superconducting and insulating $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$. *Phys. Rev. B* **63**, 172501–172505 (2001).
14. Katano, S., Sato, M., Yamada, K., Suzuki, T. & Fukase, T. Enhancement of static antiferromagnetic correlations by magnetic field in a superconductor $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ with $x = 0.12$. *Phys. Rev. B* **62**, R14677–R14680 (2000).
15. Lake, B. *et al.* Spins in the vortices of a high temperature superconductor. *Science* **291**, 832–834 (2001).
16. Nohara, M. *et al.* Quasiparticle density of states of clean and dirty d -wave superconductors: mixed-state specific heat of $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$. *J. Phys. Soc. Jpn* **69**, 1602–1605 (2001).
17. Lake, B. *et al.* Antiferromagnetic vortex state in a high-temperature superconductor. Preprint cond-mat/0104026 at (<http://xxx.lanl.gov>) (2001).
18. Savici, A. T. *et al.* Static magnetism in superconducting stage-4 $\text{La}_2\text{CuO}_{4+y}$ ($y = 0.12$). *Physica B* **289–290**, 338–342 (2000).
19. Aeppli, G. *et al.* Magnetic order and fluctuations in superconducting UPt₃. *Phys. Rev. Lett.* **60**, 615–618 (1988).
20. Arovas, D. P., Berlinsky, A. J., Kallin, C. & Zhang, S.-C. Superconducting vortex with antiferromagnetic core. *Phys. Rev. Lett.* **79**, 2871–2874 (1997).
21. Hedegård, P. Magnetic vortices in high-temperature superconductors. Preprint cond-mat/0102070 at (<http://xxx.lanl.gov>) (2001).
22. Hu, J.-P. & Zhang, S.-C. Theory of static and dynamic antiferromagnetic vortices in LSCO superconductor. Preprint cond-mat/0108273 at (<http://xxx.lanl.gov>) (2001).
23. Demler, E., Sachdev, S. & Zhang, Y. Spin ordering quantum transitions of superconductors in a magnetic field. *Phys. Rev. Lett.* **87**, 067202–067205 (2001).
24. Aeppli, G. *et al.* Nearly singular magnetic fluctuations in the normal state of a high- T_c superconductor. *Science* **278**, 432–435 (1997).
25. Boeinger, G. S. *et al.* Insulator-to-metal crossover in the normal state of $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ near optimal doping. *Phys. Rev. Lett.* **77**, 5417–5420 (1996).
26. Kittel, C. *Introduction to Solid State Physics* 6th edn, 426–428 (Wiley & Sons, New York, 1986).

Acknowledgements

We thank P. Dai, P. Hedegård, S. Kivelson, H. Mook, J. Zaanen, S. Sachdev and S.-C. Zhang for discussions. Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the US Department of Energy. H.M.R. holds a Marie Curie fellowship funded by the European Community.

Correspondence and requests for materials should be addressed to B.L. (e-mail: bella.lake@physics.ox.ac.uk).

Shear instabilities in granular flows

David J. Goldfarb[†], Benjamin J. Glasser^{*} & Troy Shinbrot^{*}

^{*} Department of Chemical and Biochemical Engineering, Rutgers University, Piscataway, New Jersey 08854-8058, USA

Unstable waves have been long studied in fluid shear layers^{1–3}. These waves affect transport in the atmosphere and oceans, in addition to slipstream stability behind ships, aeroplanes and heat-transfer devices. Corresponding instabilities in granular flows have not been previously documented⁴, despite the importance of these flows in geophysical and industrial systems^{5–7}. Here we report that breaking waves can form at the interface between two streams of identical grains flowing on an inclined plane downstream of a splitter plate. Changes in either the shear rate or the angle of incline cause such waves to appear abruptly. We analyse a granular flow model that agrees qualitatively with our experimental data; the model suggests that the waves result from competition between shear and extensional strains in the flowing granular bed. We propose a dimensionless shear number that governs the transition between steady and wavy flows.

Shear transmission in granular flow is intrinsic to systems ranging from geophysical flows to industrial processing, and peculiarities in granular shear response have consequences that can be dramatic or commonplace. More spectacular examples include

[†] Present address: Schering-Plough Research Institute, Kenilworth, New Jersey 07033, USA.

catastrophic events such as landslides⁵ and earthquakes^{8,9}, while more everyday examples include processing of products ranging from pharmaceuticals to catalysts, where jamming¹⁰, segregation^{7,11} and other endemic pitfalls reportedly reduce prevailing manufacturing efficiencies to 60% of design ratings⁷. Studies of shear in granular materials date back at least to Bagnold's pioneering research on desert sand transport more than half a century ago⁶. More recently, research into shear in well-instrumented cells^{12–15}, chutes^{11,16–19} and other systems^{10,20} has attempted to clarify the mechanisms of granular response to controlled stresses. But granular beds often display solid-like and fluid-like qualities side-by-side within a single experiment, and generally applicable rules underlying the response of a granular bed to applied shear remain elusive.

In fluids research, a geometry used to probe the responses of liquids and gases to shear is flow downstream of a splitter plate³. This flow can exhibit the Kelvin–Helmholtz instability, leading to transitions between steady, oscillatory, and turbulent states that have been productively investigated for over a century¹. However, shearing flows downstream of a splitter plate do not seem to have been reported in the granular research literature, although these flows show behaviours that are as rich and enlightening as those of fluids.

In the large pictures in Fig. 1, we show interfaces between a rapid, white, and a slower, black, stream of grains flowing down an inclined chute. The grains in each stream are coloured but otherwise identical (see Fig. 1 legend). The steady interface shown in Fig. 1a is produced at steep chute angles and high shear rates (defined below), and the other states are generated by decreasing the chute angle. Similar transitions are produced by increasing the rate of shear between the two streams. We regulated the shear flow by one of two methods: by raising one of the hopper outlets, the inlet speed was increased, or by changing the size of one of the hopper openings, the volumetric flow rate was varied. We found that raising a hopper outlet degraded reproducibility by increasing the effects of granular spray²¹. Consequently, in the results reported here we controlled the hopper opening diameters, keeping the heights of the hoppers fixed and close to the chute.

The degree of waviness was characterized by measuring the peak-to-peak amplitude of the waves from digitized video footage of the chute. This amplitude was normalized with respect to the chute width and plotted in Fig. 2b as a function of chute angle, θ . Below a critical angle of $\theta_c \approx 23^\circ$, a qualitative growth in both the mean wave amplitude and its standard deviation occurs. The increase in the mean reflects an abrupt appearance of interfacial waves below θ_c , and the growth in the standard deviation is associated with the fact that both waves and tongues are non-periodic and of variable amplitude. In experiments carried out using a glass-bottomed chute, we observed wavy instabilities from below as well as above the bed.

To identify relevant physical effects, we modified the experiment in a number of ways. Plexiglas and glass chute surfaces were used, as were plexiglas and cardboard hopper materials; these choices increased the experimental variability and are not reported here. Likewise, experiments were performed using unsieved sand, but it was found that by using sieved sand we substantially reduced the intermittency of the waves and tongues. Monodisperse spherical glass beads in a variety of sizes were also tested, but these never produced interfacial waves under any conditions. Roughening either side wall by attaching sandpaper had a minimal effect on the flow, indicating that the instability is not strongly affected by side boundary conditions.

We quantitatively probed details of the instability by taking high-speed video footage of the surface of the bed and evaluating the velocity field using particle image velocimetry²² (PIV), as shown in Fig. 2a. We scrutinized both the average and the individual PIV velocity fields, but we saw no convincing indications of vorticity—known to characterize shear instabilities of fluids—either in the

wavy or in the breaking wave state. Moreover, although Fig. 2a was obtained at $\theta = 20^\circ$, where the waves are fully developed, PIV plots at higher angles were qualitatively indistinguishable. Quantitatively, however, we found a significant difference in the rates of elongational strain seen in the steady and the wavy states. This is shown in Fig. 3, where we plot mean downstream velocities as a function of transverse distance across the chute both near the top and near the bottom of the chute for two representative cases. First, we show the velocity profile at $\theta = 20^\circ$, where interfacial waves and tongues are prevalent (compare Fig. 1c); second, we show the profile at $\theta = 28^\circ$, where waves are minimal and tongues are never seen (compare Fig. 1a). From Fig. 3, we see that at $\theta = 20^\circ$, grains accelerate only slightly, about 0.3 m s^{-1} from top to bottom of the chute, while at $\theta = 28^\circ$, grains accelerate more than twice as much in the same distance. However, the shear rates (that is, the slopes of these plots near the left of the chute) for all four cases are comparable, differing by less than 20% from one another. We will return to this point below.

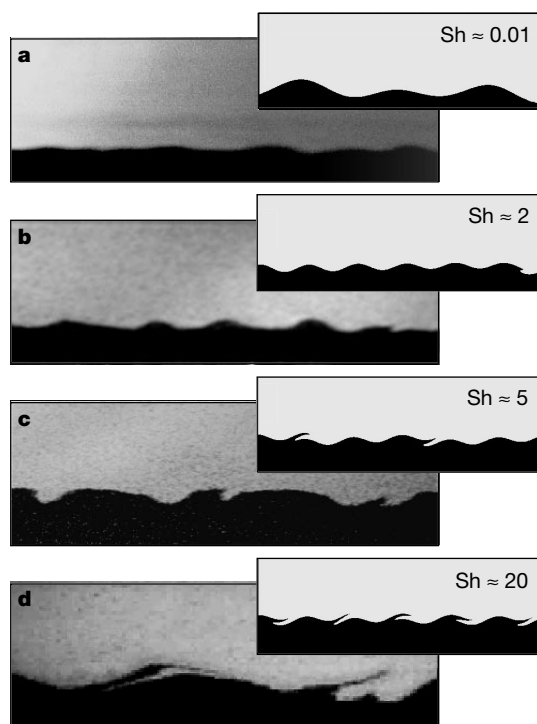


Figure 1 Instability of interface between shearing granular streams. Views from above of a sheared granular chute flow: flow is from left to right. Growth of the wavy instability occurs as the inclination of the chute to the horizontal, θ , decreases: **a**, $\theta = 28^\circ \pm 0.2^\circ$, **b**, $\theta = 22^\circ \pm 0.2^\circ$, and **c**, $\theta = 19.6^\circ \pm 0.2^\circ$. Picture **d** was taken at $\theta = 24^\circ \pm 0.2^\circ$, but where the extensional rate of flow in the granular bed was artificially reduced (see text). The two streams of sand were fed by independent hoppers, and the mass flow rates of the white and black sand were maintained respectively at 295 ± 2 and $70 \pm 3 \text{ g s}^{-1}$. Flow rates were determined by discharging either hopper onto a digital scale, and interrogating the scale output 30 times a second. The snapshots capture the entire width of the chute, where the left edge of each snapshot is about 30 cm downstream of the splitter plate. The splitter plate is placed midway across the chute, but the higher mass flow of the white sand compresses the black region, and the streams equilibrate to nearly the same height. Both black and white grains are 'art sand'; in snapshots **a–c** the sand has been sieved to below $350 \mu\text{m}$ diameter; in **d**, unsieved sand was used (see text). The chute consists of an inclined sheet of plexiglas with plexiglas side walls. The working surface of the chute was 7.6 cm wide and 107 cm long, and was fitted with an aluminium insert. Insets: comparison model interfacial waves for shearing and elongating bed after 100 natural time units at several shear numbers (Sh), as defined in the text to be the ratio between shear and elongational strain rates. Movies of the shearing instability are available at <http://sol.rutgers.edu/~bglasser/Shear>.

Evidently, there is a transition between a steady and a wavy state of the interface between two shearing granular streams. Analysis of this and other granular problems is hindered by a lack of generally applicable equations of motion^{4,23}. Nevertheless, in this problem there are two well-documented phenomenological facts that can be used to analyse the flow.

The first pertinent fact is that Mach numbers of granular flows, expressed as a ratio of mean to fluctuational velocities, commonly exceed one^{24,25}, and so these flows produce travelling shocks²⁴. In chute flows, these shocks have been reported in the form of density (in one or two dimensions) or height (in three dimensions) waves whose speed depends on mean bed velocity^{4,16,24}. Thus at the interface between shearing granular streams, two sets of shocks must travel with different speeds and cannot, therefore, remain aligned everywhere. So at the interface between these streams, there must of necessity be a differential in density or height that itself travels up- or down-stream^{16–18,25,26}.

In addition to this transverse gradient, grains are advected in the streamwise direction by an imposed shear. We can construct a computational model that subjects an interface to these two fields.

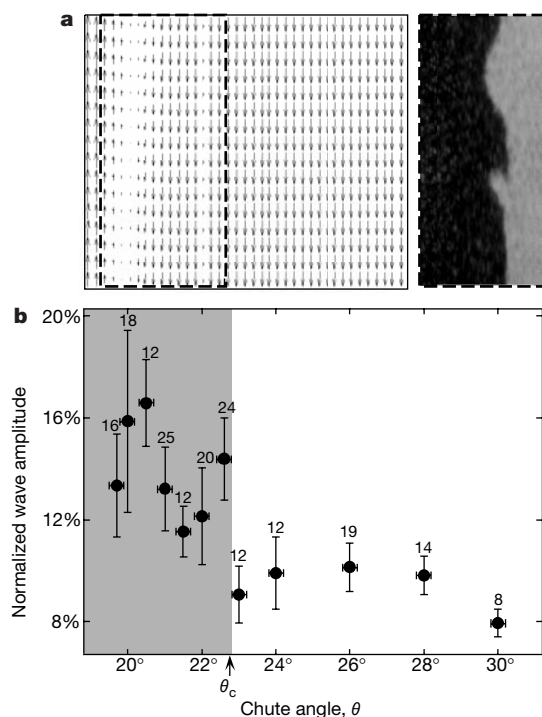


Figure 2 Onset of shear instability. **a**, Representative PIV velocity field of shearing granular chute flow at 20° inclination. The velocity vectors shown are averaged over 100 image pairs from video footage captured at 500 frames per second. The mean velocity has been subtracted to expose the shear. We also show a section of the interface, identified in the plot within a dashed rectangle. **b**, Mean peak-to-peak amplitudes of waves normalized to the chute width as a function of chute inclination. Critical inclination, θ_c , below which transverse waves appear, is about 23° for our experiments. Below an inclination of $\theta \approx 19.6^\circ$, flow cannot be sustained. Vertical error bars represent standard deviations from means taken over between 8 and 25 trials; horizontal bars are associated with uncertainties in angular measurement (obtained using a digital level). Numbers of trials are shown above data: more trials were performed at larger wave amplitudes to refine the mean value recorded. All experiments were performed in a controlled environmental chamber, set at 20 °C and 65% relative humidity. To minimize static charging, earthed aluminium flashing was fitted into the bottom of the chute, the hoppers were made of galvanized steel, and all chute surfaces were cleaned before each run with anti-static spray. Despite these safeguards, the waves and tongues that we report were persistently aperiodic, and multiple trials under carefully controlled conditions were necessary to produce reliable statistics.

We take the simplest possible ansatz, namely that a particle at position (X, Y) travels unencumbered with streamwise and spanwise velocities given by equations (1) and (2) respectively:

$$v_x = af(Y) + b \tanh(k_y Y) \quad (1)$$

$$v_y = c \sin(k_x X - \omega T) \quad (2)$$

Here X is the streamwise and Y the spanwise coordinate, T denotes time, and the remaining quantities are constant parameters. For simplicity, we apply these velocities on a domain that is periodic at $X = \pm 1$ and unbounded in Y . This model represents neither the only nor necessarily the most accurate analysis of the phenomenon, but it is analytically simple and it captures essential physics present in our problem. The function $f(Y)$ only serves to advect material downstream, and is chosen on the basis of the best available data^{12,15,20,27} to be of the form:

$$f(Y) = [\cosh(\kappa\pi) - \cosh(\kappa Y)] \quad (3)$$

where κ governs the thickness of the boundary layer at the edges of the chute. Setting $\kappa = 5$ approximately matches our measured velocity fields (Fig. 3); comparison simulations (not shown) reveal that results vary only quantitatively with κ . In the inset to Fig. 3, we apply these prescribed velocity fields to an initially straight interface, $Y = 0$, and plot the interface's evolution for a representative set of parameters ($a = 0.0005$, $b = 0.3$, $c = 0.1$, $k_x = 3$, $k_y = 6$, $\kappa = 5$, $\omega = 0.1$). Simply integrating the prescribed velocities causes the interface to evolve first into a wavy curve, whose peaks subsequently become stretched by the shearing field into tongue-like projections.

Kinematically, this is a predictable consequence of equations (1)–(3) under steady conditions: in a frame moving downstream with speed ω/k_x , equation (2) has the steady solution: $Y = -(c/\omega)\cos(\varphi)$, where $\varphi = k_x X - \omega T$. Steady shear causes this wave to steepen and break, and indeed breaking waves are

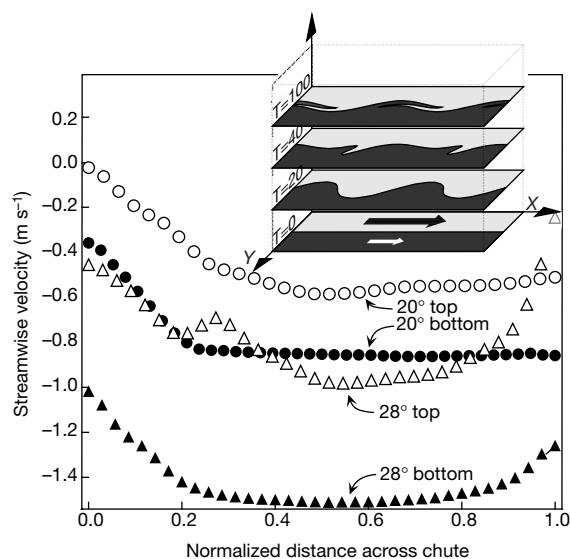


Figure 3 Velocity field in shearing streams. Main plot: streamwise velocity at top (20 cm downstream of splitter plate) and bottom (70 cm downstream) of chute averaged from PIV measurements at two chute inclinations—20°, in the fully developed wavy regime, and 28°, in the stable flow regime. The shear rates near the left of the figure are comparable in these two plots, but the rate of acceleration from top to bottom of the chute differs substantially in the two cases. Inset: evolution of an initially linear interface between a fast (light) and a slow (dark) stream of material subject to model velocity field given by equations (1)–(3). In sequential time slices shown in this plot, transverse interfacial waves emerge and then break to form tongue-like projections. The simulation shown here is periodic at the left and right of each timeslice.

observed in our simulations for a broad range of parameter choices.

This brings us to the second well-documented fact: under non-steady conditions, grains are continually accelerated by gravity, so a granular bed elongates as grains travel downstream^{18,24}. This can stabilize the breaking waves, because elongation stretches disturbances along the streamwise direction, thereby counteracting the steepening influence of shear on emerging waves. The competition between elongation and shear can be directly analysed: if the elongational strain rate is $\dot{\gamma}_e$ and the concurrent shear strain rate is $\dot{\gamma}_s$, then a leading edge of a wave with slope $\delta Y/\delta X$ evolves in time δT according to:

$$\frac{\delta Y}{\delta X} \rightarrow \frac{\delta Y}{\delta X + \delta X \dot{\gamma}_e \delta T - \delta Y \dot{\gamma}_s \delta T} \quad (4)$$

Therefore a disturbance steepens or flattens according to whether the dimensionless ratio between the two last terms in the denominator is more or less than unity. Thus we define a dimensionless shear number²⁸, Sh, that governs the stability of this problem:

$$\text{Sh} = \frac{\delta Y}{\delta X} \frac{\dot{\gamma}_s}{\dot{\gamma}_e} \quad (5)$$

where we have included a generalized dispersion term, $\dot{\epsilon}$, to account for other—for example, noise related—effects that can stabilize even the steady case, $\dot{\gamma}_e = 0$, for sufficiently slow shear. The presence of the slope as a prefactor indicates that this is a finite-amplitude instability that only grows if a disturbance has slope above a critical value.

We can examine the viability of this analysis computationally by subjecting the model defined by equations (1)–(3) to uniform streamwise elongation. Doing so at representative parameter values, we obtain the states shown in the insets to Fig. 1 for several values of Sh. Beyond this qualitative agreement between experiment and model computations, we performed two direct tests of the significance of Sh on interfacial wave formation. First, we evaluated Sh from our PIV data (taking $\dot{\epsilon} = 0$), and as anticipated, we found that Sh grew by 60% as θ was lowered from 20° to 28°. (The shear itself varied by less than 20%.) Second, we roughened the last 10 cm of the chute bottom, thus artificially retarding downstream particles and decreasing the elongational strain. This produced an augmentation of the breaking waves: the picture shown in Fig. 1d was obtained in such an experiment—in fact, the sand used in this experiment was not sieved, and seldom formed tongues without artificially reducing the elongation in this way.

These results show that, under controlled conditions, the instability reported here is clear-cut and reproducible, but the precise values of shear rate and inclination at which the instability appears depend sensitively on material and process parameters. For example, our inlet streams are fed by hoppers, which are known to produce stochastic flow rates²⁹ that could influence the instability; additionally, convection and other complicating effects¹⁷ have been recorded in chute flows¹¹. Moreover, waves and tongues are most accentuated at low elongational strains, which for our apparatus corresponds to a narrow parameter range near where the flow stalls. Two points therefore deserve emphasis. First, the instability that we report is seen only in experiments with shear applied across the splitter plate, and not in comparison experiments performed without shear, but in which the mean flow rate was varied. Second, in separate experiments we intentionally vibrated the splitter plate to study the convection of disturbances, but we found that this had little effect on the growth of waves downstream. These control experiments indicate that the reported instability depends intrinsically on shear between granular streams, and is not caused by extrinsic fluctuations.

Unlike comparable fluids instabilities, the granular waves that we report here do not appear to be associated with localized vorticity;

rather, the phenomenon seems to be consistent with a competition between shear and elongation of the granular bed in the presence of differential travelling shocks. Because both shocks and shear are commonly seen in granular flows, it seems plausible that this instability may be present in many physical and engineering problems. In geophysical debris flows, for example, wide lateral variations in sedimentation⁵, erosion³⁰ and property damage²⁸ have been reported, all of which intrinsically involve strong shear gradients. Likewise in industrial powder flows, the response to shear gradients is central to the accurate prediction of mixing and transport rates. As has proven to be the case in studies of fluids, detailed investigation of underlying instabilities may prove fruitful for both the improved understanding of the mechanism of shear transmission and the predictive analysis of the behaviours of granular flows of practical importance. □

Received 12 July; accepted 23 November 2001.

- Lamb, H. *Hydrodynamics* (Cambridge Univ. Press, Cambridge, 1932).
- Chandrasekhar, S. *Hydrodynamic and Hydromagnetic Stability* (Clarendon, Oxford, 1961).
- Thorpe, S. A. Experiments on the instability of stratified shear flows: immiscible fluids. *J. Fluid Mech.* **39**, 25–48 (1969).
- Goddard, J. D. & Alam, M. Shear-flow and material instabilities in particulate suspensions and granular media. *Particulate Sci. Technol.* **17**, 69–96 (1999).
- Denlinger, R. P. & Iverson, R. M. Flow of variably fluidized granular masses across three-dimensional terrain. 2. Numerical predictions and experimental tests. *J. Geophys. Res.* **106**, 553–566 (2001).
- Bagnold, R. A. The movement of desert sand. *Proc. R. Soc. Lond. A* **157**, 594–620 (2001).
- Shinbrot, T. & Muzzio, F. J. Nonequilibrium patterns in granular mixing and segregation. *Phys. Today* **25–30** (March, 2000).
- Melosh, H. J. Dynamical weakening of faults by acoustic fluidization. *Nature* **379**, 601–606 (1996).
- Scott, D. R. Seismicity and stress rotation in a granular model of the brittle crust. *Nature* **381**, 592–595 (1996).
- Liu, A. J. & Nagel, S. R. Nonlinear dynamics—Jamming is not just cool any more. *Nature* **396**, 21–22 (1998).
- Pouliquen, O., Delour, J. & Savage, S. B. Fingering in granular flows. *Nature* **386**, 816–817 (1997).
- Losert, W., Bocquet, L., Lubensky, T. C. & Gollub, J. P. Particle dynamics in sheared granular matter. *Phys. Rev. Lett.* **85**, 1428–1431 (2000).
- Menon, N. & Durian, D. J. Diffusing-wave spectroscopy of dynamics in a three-dimensional granular flow. *Science* **275**, 1920 (1997).
- Howell, D. & Behringer, R. P. Fluctuations in a 2D granular Couette experiment: A critical transition. *Phys. Rev. Lett.* **82**, 5241–5244 (1999).
- Mueh, D. M. et al. Signatures of granular microstructure in dense shear flows. *Nature* **406**, 385–389 (2000).
- Louge, M. Y. & Keast, S. C. On dense granular flows down flat frictional inclines. *Phys. Fluids* **13**, 1213–1233 (2001).
- Forterre, Y. & Pouliquen, O. Longitudinal vortices in granular flows. *Phys. Rev. Lett.* **86**, 5886–5889 (2001).
- Hanes, D. M. & Walton, O. R. Simulations and physical measurements of glass spheres flowing down a bumpy incline. *Powder Technol.* **109**, 133–144 (2000).
- Savage, S. B. & Hutter, K. The motion of a finite mass of granular material down a rough incline. *J. Fluid Mech.* **199**, 177–215 (1989).
- Pouliquen, O. & Gutfrand, R. Stress fluctuations and shear zones in quasistatic granular chute flows. *Phys. Rev. E* **53**, 552–561 (1996).
- Davies, C. E., Weir, G. & McGavin, P. Behaviour of a dense stream of non-cohesive particles impacting on an inclined plate. *Powder Technol.* **106**, 1–6 (1999).
- Lueptow, R. M., Akonur, A. & Shinbrot, T. PIV for granular flows. *Exp. Fluids* **28**, 183–186 (2000).
- Kadanoff, L. P. Built upon sand: Theoretical ideas inspired by granular flows. *Rev. Mod. Phys.* **71**, 435–444 (1999).
- Johnson, P. C., Nott, P. & Jackson, R. Frictional-collisional equations of motion for particulate flows and their application to chutes. *J. Fluid Mech.* **210**, 501–535 (1990).
- Reicha, E. C., Bizon, C., Shattuck, M. D. & Swinney, H. L. Shocks in supersonic sand. *Phys. Rev. Lett.* **88**, 014302 (2002).
- Wang, C.-H., Jackson, R. & Sundaresan, S. Instabilities of fully developed rapid flow of a granular material in a channel. *J. Fluid Mech.* **342**, 179–197 (1997).
- Natarajan, V. V. R., Hunt, M. L. & Taylor, E. D. Local measurements of velocity fluctuations and diffusion coefficients for a granular material flow. *J. Fluid Mech.* **304**, 1–25 (1995).
- Iverson, R. M. The physics of debris flows. *Rev. Geophys.* **35**, 245–296 (1997).
- Aider, J. L., Sommer, N., Raafat, T. & Hulin, J. P. Experimental study of a granular flow in a vertical pipe: a spatiotemporal analysis. *Phys. Rev. E* **59**, 778–786 (1999).
- Chan, K. & Rothman, D. H. Coupled length scales in eroding landscapes. *Phys. Rev. E* **63**, 55102–55105 (2001).

Acknowledgements

We thank A. Hasan, J. Leung, E. Liss, K. Mehta and J. Pantina for assistance. This work was supported by the US National Science Foundation, Division of Chemical and Transport Systems, and the American Chemical Society, Petroleum Research Fund.

Correspondence and requests for materials should be addressed to B.J.G. (e-mail: bglasser@sol.rutgers.edu).

Observation of conformation-specific pathways in the photodissociation of 1-iodopropane ions

Sang Tae Park*, Sang Kyu Kim† & Myung Soo Kim*

* National Creative Research Initiative for Control of Reaction Dynamics and School of Chemistry, Seoul National University, Seoul 151-742, Korea

† Department of Chemistry, Inha University, Incheon 402-751, Korea

Many molecules can rotate freely around single bonds and thereby interconvert between different conformations, such as *gauche* and *anti* 1,2-disubstituted ethane, a classic example of conformational isomerism^{1–3}. Even though rotation occurs rapidly at room temperature, the product selectivity seen in some reactions has been explained by conformation-dependent reaction mechanisms^{4,5}: if reactant molecules differing only in their conformation are located at different positions on the reaction path, they may undergo different reactions. But a direct verification of this effect is difficult, because the energy barrier separating conformational isomers is so low that under ambient conditions reactants with more than one conformation will be present¹. But by using temperatures low enough to suppress the interconversion between different conformations, *gauche*-1-iodopropane ions and *anti*-1-iodopropane ions have been selectively generated⁶. Here we show that the kinetic energy released during the photodissociation of 1-iodopropane ions depends strongly on the conformation of the ions. Thermodynamic arguments and *ab initio* calculations indicate that this difference in kinetic energy release results from differences in the reaction mechanism, with *gauche*-1-iodopropane ions forming 2-propyl ions and *anti*-1-iodopropane ions forming protonated cyclopropane ions. These findings suggest that the well-known concept of conformation selection forms the basis of a simple scheme for reaction control, thus providing in some cases an attractive alternative for more involved schemes that utilize the phase and pulse shape of laser beams to control chemical reactions^{7–10}.

We have previously⁶ used mass-analysed threshold ionization¹¹ (MATI), a technique for generating state-selected molecular ions, to produce 1-iodopropane ions ($1\text{-C}_3\text{H}_7\text{I}^+$) with coherent vacuum-ultraviolet¹² (VUV) radiation (Fig. 1). The resonant nature of MATI and the very low temperature (rotational temperature ~ 35 K) of the molecular beam ensure that $1\text{-C}_3\text{H}_7\text{I}^+$ is in its ground vibronic state, with hardly any internal energy that could drive the interconversion between the *gauche* and *anti* forms of the ion. These two forms should therefore be considered as isomers, rather than conformers. Interconversion between the *gauche* and *anti* forms of $1\text{-C}_3\text{H}_7\text{I}^+$ can also be induced by collisions, but the very low pressure ($< 10^{-7}$ torr) in our experimental system precludes this process. Hence, we are able to produce ion beams containing only *gauche*- $1\text{-C}_3\text{H}_7\text{I}^+$ or only *anti*- $1\text{-C}_3\text{H}_7\text{I}^+$, by selectively tuning the wavelength of the ionizing VUV radiation (133.94 and 133.72 nm for the *gauche* and *anti* forms, respectively).

However, conformation-selective formation of ion beams does not guarantee the observation of a conformation-selective reaction, because the activation energy needed to initiate a reaction may also induce the interconversion between different conformations. We therefore used photoinduced excitation of the molecular ions to a repulsive electronic state, which ensures that subsequent dissociation occurs essentially immediately. This approach should prevent complications arising from interconversion between the conformations.

In our experiments, selective generation of either *gauche* or *anti* $\text{C}_3\text{H}_7\text{I}^+$ ions by a VUV laser pulse was followed about 100 ns later by

a second laser pulse in the visible wavelength region (480–700 nm), which excited the ions to their first excited electronic state¹³—according to our observations (see below), this excited state is repulsive with respect to the C–I bond. The C_3H_7^+ fragment ion produced in this photodissociation event was then accelerated orthogonal to the molecular beam/VUV direction, and detected by time-of-flight (TOF) mass spectrometry.

The TOF profiles of C_3H_7^+ changed with the polarization of the photodissociation laser, indicating that dissociation via the first excited state occurs very rapidly and hence anisotropically, as was intended. The anisotropy effect can be eliminated by recording a TOF profile at the laser polarization angle of 54.7° : the broadening of the profile recorded at this angle is then solely due to the centre-of-mass kinetic energy of the fragments, the kinetic energy release KER (ref. 14). Figure 2a shows the TOF profiles of C_3H_7^+ generated from *gauche* and *anti* $1\text{-C}_3\text{H}_7\text{I}^+$ by using a laser pulse with a wavelength of 607 nm and a polarization angle of 54.7° . The profile generated from the *gauche* isomer is significantly broader than that from the *anti* form, indicating that the photodissociation of the *gauche* form results in the release of a larger amount of kinetic energy. Figure 2b illustrates the KER distribution obtained from the TOF profiles with a method similar to that in ref. 15. These distributions yield average values of KER, $\langle T \rangle$, of 0.096 eV and 0.042 eV for the photodissociation at 607 nm of the *gauche* and *anti* forms of $\text{C}_3\text{H}_7\text{I}^+$, respectively.

We performed further photodissociation experiments using dissociating laser pulses with different wavelengths in the range 480–700 nm. Figure 2c shows the resulting plot of $\langle T \rangle$ obtained from these data as a function of photon energy. The figure also includes the KER measured by Brand *et al.*¹⁶ for the same reaction, using a photoionization technique that does not allow for state and conformation selections (see Fig. 2c). The data show that even when $\text{C}_3\text{H}_7\text{I}^+$ has an internal energy of only ~ 1.1 eV (so dissociation can occur only in the ground electronic state), the KER observed by Brand *et al.* is much larger than the KER we observe near the dissociation threshold (700 nm, 1.77 eV). This difference indicates that the dissociation events we probe do not involve the ground electronic state of $\text{C}_3\text{H}_7\text{I}^+$, and that our experimental set-up thus performs as planned.

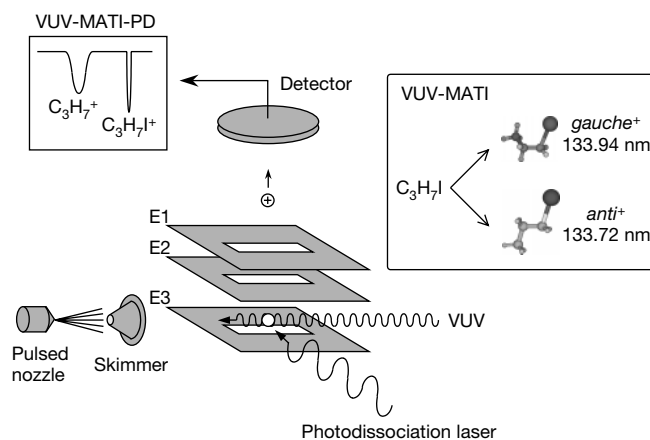


Figure 1 Experimental scheme. Vacuum-ultraviolet radiation (VUV) at a wavelength of ~ 134 nm is generated in a Kr cell by combining and focusing two laser beams at 212 and ~ 510 nm. This VUV is overlapped with the molecular beam in a collinear fashion. Electrode E3 is kept at -0.5 V, and other electrodes grounded at the time of the VUV pulse. After $20 \mu\text{s}$ delay, 930 and 1,200 V are applied to electrodes E2 and E3, respectively, for the pulsed-field ionization and acceleration. The *gauche* and *anti* isomers of $1\text{-C}_3\text{H}_7\text{I}^+$ are generated selectively by VUV-mass-analysed threshold ionization (VUV-MATI) at 133.94 and 133.72 nm, respectively (inset). About 100 ns after the ionization, the sample is irradiated by a second laser in the visible to induce photodissociation (PD).

The values obtained for the photodissociation of the *gauche* and *anti* forms of $C_3H_7I^+$ decrease almost linearly with photon energy and with similar slopes, but the absolute values obtained for the *gauche* isomer are always noticeably larger than those obtained for the *anti* form. Assuming that $\langle T \rangle$ depends linearly on the photon energy, an extrapolation of the data to $\langle T \rangle = 0$ yields dissociation threshold energies of 1.6 eV and 1.3 eV for *anti* $C_3H_7I^+$ and *gauche* $C_3H_7I^+$, respectively.

When trying to explain the conformation-specific differences in the photodissociation behaviour that we observe, we first note that the chemical composition of the reactants and products involved are identical. Assuming the reaction paths for the reactions involving the *gauche* and *anti* forms of $C_3H_7I^+$ are the same, the difference in the KER would have to be due to a significant difference in the internal energy of the photoexcited *gauche* and *anti* isomers. This, however, can be ruled out: the internal energy of the *anti* form exceeds that of the *gauche* form by ~ 0.015 eV (ref. 6), yet the latter releases more kinetic energy when dissociating. The significant difference in KER (Fig. 2b) and the difference in the dissociation threshold energies estimated from Fig. 2c thus seem to result from differences in the dissociation dynamics of the two $C_3H_7I^+$ forms (see, for example, ref. 17). The fact that the KER distributions for

gauche and *anti* $C_3H_7I^+$ (Fig. 2b) are conspicuously different and do not appear to be the superposition of two components also suggests that photoexcitation does not induce the interconversion of the two $C_3H_7I^+$ conformations, but that they move along different reaction paths. We can also exclude a photodissociation reaction mechanism that involves internal conversion to the ground state followed by dissociation. Such a mechanism would not be expected to lead to the conformation-specific differences we have observed in our data. This is because internal conversion to the ground state would yield $C_3H_7I^+$ species with a large amount of internal energy, and thus able to rapidly interconvert between the two different conformational states. These considerations, taken together, suggest that the reactions we have probed occur rapidly and anisotropically in an excited state that is repulsive, or the repulsive region of which is being probed in the present experiment.

To explain our observations, we consider the possible products that may be formed in our system. Simple C–I bond cleavage of 1- $C_3H_7I^+$ would generate 1- $C_3H_7^+$ and I, but 1- $C_3H_7^+$ is not a stable gas-phase species¹⁸. However, the 2-propyl cation (2- $C_3H_7^+$) and protonated cyclopropane (cyclo- $C_3H_7^+$) are stable species that are generated during the photodissociation of 1- $C_3H_7I^+$ (refs 19, 20). Consideration of the energies involved in the photodissociation process suggest that the iodine atom will be generated as one of the two possible ground-state spin–orbit doublets, $^2P_{3/2}$ or $^2P_{1/2}$. Plausible product combinations are therefore (1) 2- $C_3H_7^+$ + I ($^2P_{3/2}$), (2) 2- $C_3H_7^+$ + I ($^2P_{1/2}$), (3) cyclo- $C_3H_7^+$ + I ($^2P_{3/2}$), and (4) cyclo- $C_3H_7^+$ + I ($^2P_{1/2}$). The fact that the KER distributions in Fig. 2b do not appear as a superposition of two or more components indicates that the photoexcited *gauche* and *anti* 1- $C_3H_7I^+$ each dissociate to give only one of four possible product combinations. The estimated reaction enthalpies for the corresponding reactions are listed in Table 1. Our estimated dissociation threshold energies of 1.3 eV for *gauche*-1- $C_3H_7I^+$ and 1.6 eV for *anti*-1- $C_3H_7I^+$ suggest that the dissociation of these isomers involves the respective product channels (2) and (4), as these are associated with reaction enthalpies of 1.40 and 1.72 eV at 0 K, respectively (Table 1). That is, thermochemical analysis suggests that dissociation of the *gauche* isomer produces 2- $C_3H_7^+$ while the *anti* form generates the less-stable cyclo- $C_3H_7^+$ ion.

The potential energy diagram along the two reaction paths of 1- $C_3H_7^+$ in the first excited electronic state ($n \leftarrow \sigma$ (C–I) character), obtained from *ab initio* calculations at the single excitation configuration interaction²¹ (CIS) level, is shown in Fig. 3. In accord with our earlier considerations, the calculated excited state is repulsive with respect to the C–I bond. The calculations show that as *gauche*-1- $C_3H_7I^+$ moves along the reaction path, the C–I bond is lengthened. About halfway along the reaction path, the hydrogen atom on C(2) that occupies the *anti* position with respect to the iodine atom

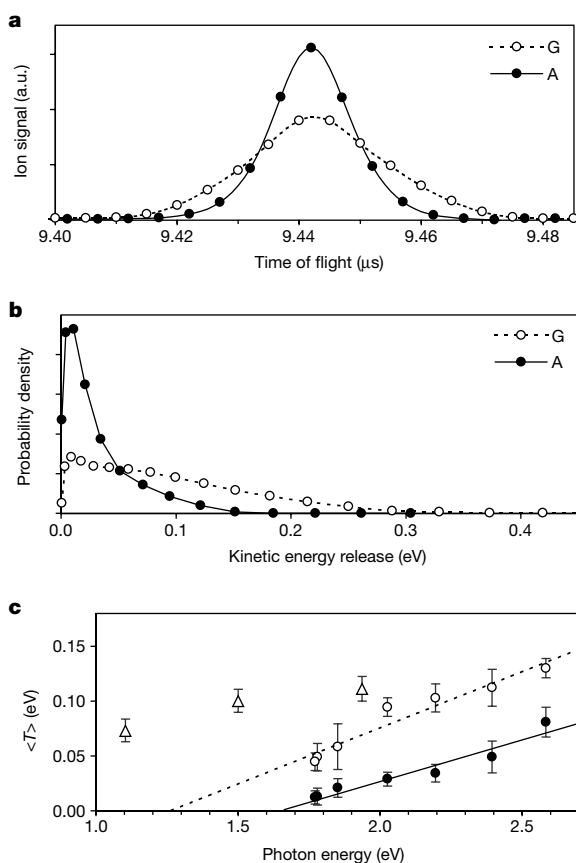


Figure 2 Time-of-flight profiles and kinetic energy releases. **a**, Time-of-flight profiles of $C_3H_7^+$ generated by photodissociation of the *gauche* (G; open circles) and *anti* (A; filled circles) isomers of 1- $C_3H_7I^+$ with 54.7° polarization of the photodissociation laser at 607 nm. Averages over 2,500 laser shots. **b**, Kinetic energy release (KER) distributions determined from **a**, *gauche* (open circles) and *anti* (filled circles) isomers. **c**, Average KER, $\langle T \rangle$, versus photon energy in the 480–700 nm range, for *gauche* (open circles) and *anti* (filled circles) isomers. Each point represents an average over ten measurements (± 2 s.d.), which, in turn, are averages over 250 laser shots. Values of $\langle T \rangle$ from the photoionization measurement of ref. 16 are shown as open triangles. a.u., arbitrary units.

Table 1 Enthalpies of formation and reaction at 0 K

System	$\Delta H_{f,0K}^0$ (eV)*	$\Delta H_{r,0K}^0$, <i>gauche</i> (eV)†	$\Delta H_{r,0K}^0$, <i>anti</i> (eV)†
Reactants			
1- $C_3H_7I^+$ (<i>gauche</i>)	9.170 $\pm$ 0.040		
1- $C_3H_7I^+$ (<i>anti</i>)	9.186 $\pm$ 0.040		
Products			
2- $C_3H_7^+$ + I ($^2P_{3/2}$)	9.627 $\pm$ 0.039	0.457 $\pm$ 0.056	0.442 $\pm$ 0.056
2- $C_3H_7^+$ + I ($^2P_{1/2}$)	10.570 $\pm$ 0.039	1.400 $\pm$ 0.056	1.385 $\pm$ 0.056
cyclo- $C_3H_7^+$ + I ($^2P_{3/2}$)	9.958 $\pm$ 0.041	0.788 $\pm$ 0.056	0.772 $\pm$ 0.056
cyclo- $C_3H_7^+$ + I ($^2P_{1/2}$)	10.901 $\pm$ 0.041	1.730 $\pm$ 0.056	1.715 $\pm$ 0.056

For products, the enthalpy of formation at 0 K is the sum of the enthalpies for the two products. Data for the reactants and 2- $C_3H_7^+$ (8.517 eV) are evaluated using thermochemical data²². Enthalpy of formation at 0 K of cyclo- $C_3H_7^+$, 8.847 eV, is evaluated using the *ab initio* results at the G2 level. The energy difference between the two fragment ions, 0.331 eV, is close to the *ab initio* result, 0.313 eV, at the MP4/6-311G** level (ref. 20). Using the enthalpy of formation of cyclo- $C_3H_7^+$ obtained from proton affinity measurement²³, 8.864 eV, the experimental difference becomes 0.347 eV, which is in agreement with the G2 result. Enthalpies of formation at 0 K of I ($^2P_{3/2}$) and I ($^2P_{1/2}$) are²⁴ 1.1107 and 2.0534 eV, respectively.

* Enthalpy of formation at 0 K.

† Enthalpy of reaction at 0 K.

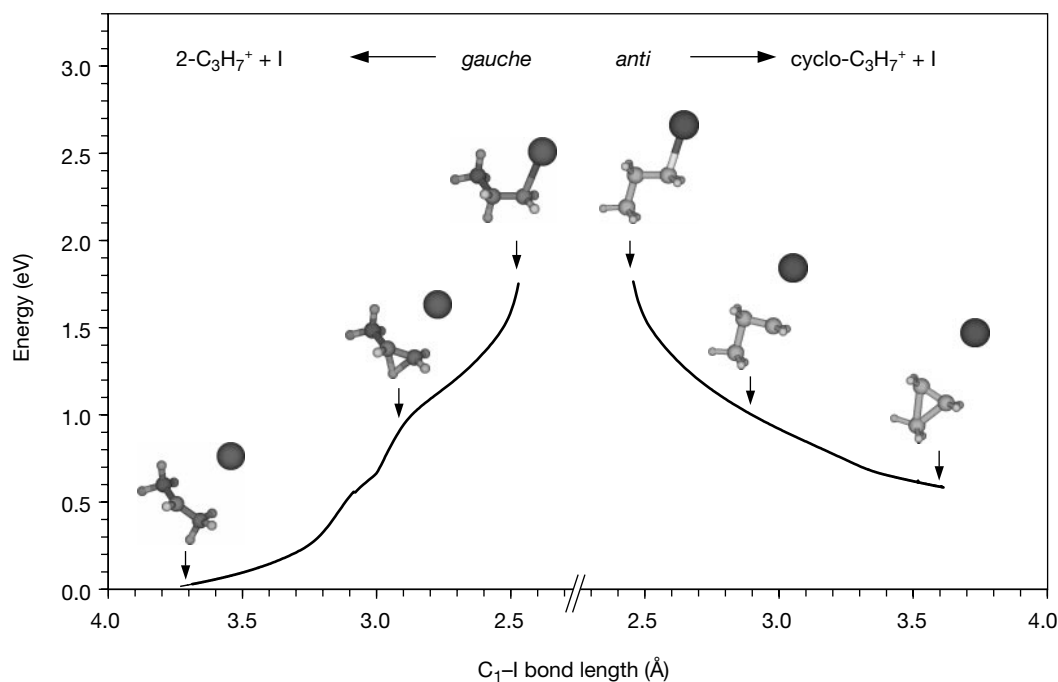


Figure 3 Potential energy along the reaction path. Changes in potential energies along the minimum energy paths for the dissociations of *gauche* and *anti* isomers in the first excited state were obtained by *ab initio* calculation at the CIS level. The 6-31G** basis set was used for carbons and hydrogens, while the LanL2DZ basis set was used for iodine. Equilibrium geometries of the *gauche* and *anti* isomers in the ground electronic state at the Hartree-Fock level were taken as the initial geometries of the photoexcited 1-C₃H₇I⁺.

Then, the energies and gradients in the first excited state corresponding to the above configurations were calculated by CIS. Finally, the minimum energy paths from these configurations were calculated by the steepest descent method. The energy of the products, 2-C₃H₇⁺ + I⁻, is taken as the zero of the energy scale. Some representative geometries are also drawn.

moves to C(1), and finally 2-C₃H₇⁺ is formed. In the case of *anti*-1-C₃H₇I⁺, movement of the methyl group occupying the *anti* position to the iodine atom accompanies the C-I bond breaking, and finally cyclo-C₃H₇⁺ is formed. In both cases, an intramolecular S_N2-type rearrangement accompanies the C-I bond breaking. The substantial changes in the C₃H₇⁺ structure during C-I bond lengthening indicates that the reaction path is curved on the multidimensional potential energy surface. This might explain why the fraction of the available energy that is released as kinetic energy is rather small, compared to typical repulsive dissociation reactions (see, for example, refs 14, 15). The presence of shallow bound regions on the potential energy surface of the reaction would provide an alternative explanation for the small fraction of available energy being released as kinetic energy. But such an explanation would be in disagreement with the *ab initio* results. Considering all the evidence presented, the most plausible explanation for our experimental observations is provided by assuming conformation-specific differences in the photodissociation mechanism of 1-C₃H₇I⁺. □

Received 24 July; accepted 23 November 2001.

1. Eliel, E. L. & Wilen, S. H. *Stereochemistry of Organic Compounds* 597–664 (Wiley, New York, 1994).
2. Kirby, A. J. *Stereochemical Effects* 24–33 (Oxford Univ. Press, New York, 1996).
3. Ramsay, O. B. *Stereochemistry* 1–42 (Heyden, London, 1981).
4. Winstein, S. & Holness, N. J. Neighboring carbon and hydrogen. XIX. t-Butylcyclohexyl derivatives. Quantitative conformational analysis. *J. Am. Chem. Soc.* **77**, 5562–5578 (1955).
5. Eliel, E. L. & Lukach, C. A. Conformational analysis. II. Esterification rates of cyclohexanols. *J. Am. Chem. Soc.* **79**, 5986–5992 (1957).
6. Park, S. T., Kim, S. K. & Kim, M. S. One-photon mass-analyzed threshold ionization spectroscopy of 1- and 2-iodopropanes in vacuum ultraviolet. *J. Chem. Phys.* **114**, 5568–5576 (2001).
7. Brumer, P. & Shapiro, M. Laser control of molecular process. *Annu. Rev. Phys. Chem.* **43**, 257–282 (1992).
8. Kosloff, R. *et al.* Wavepacket dancing—achieving chemical selectivity by shaping light-pulses. *Chem. Phys.* **139**, 201–220 (1989).
9. Zhu, L. *et al.* Coherent laser control of the product distribution obtained in the photoexcitation of HI. *Science* **270**, 77–80 (1995).

10. Assion, A. *et al.* Control of chemical reactions by feedback-optimized phase shaped femtosecond laser-pulse. *Science* **282**, 919–922 (1998).
11. Schlag, E. W. *ZEKE Spectroscopy* 112–121 (Cambridge Univ. Press, Cambridge, 1998).
12. Hilber, G., Lago, A. & Wallenstein, R. Broadly tunable vacuum-ultraviolet/extreme-ultraviolet radiation generated by resonant third-order frequency conversion in krypton. *J. Opt. Soc. Am. B* **4**, 1753–1764 (1987).
13. Kimura, K., Katsumata, S., Achiba, Y., Yamazaki, T. & Iwata, S. *Handbook of HeI Photoelectron Spectra of Fundamental Organic Molecules* 87 (Japan Scientific Societies, Tokyo, 1981).
14. Jarrold, M. F., Illies, A. J. & Bowers, M. T. Energy disposal in photodissociation from magic angle measurements with a crossed high-energy ion beam and laser beam: photodissociation dynamics of the (N₂)₂⁺ cluster in the 458–514 nm range. *J. Chem. Phys.* **81**, 214–221 (1984).
15. Kim, D. Y., Choe, J. C. & Kim, M. S. Photodissociation dynamics of the propargyl bromide molecular ion. *J. Chem. Phys.* **113**, 1714–1724 (2000).
16. Brand, W. A., Baer, T. & Klotz, C. E. Kinetic energy release distribution in the fragmentation of energy-selected iodopropane ions. *Chem. Phys.* **76**, 111–120 (1983).
17. Levens, K. *Fundamental Aspects of Organic Mass Spectrometry* 166–172 (Chemie, Weinheim, 1978).
18. Lias, S. G., Rebert, R. E. & Ausloos, P. Carbonium ions in radiation chemistry. II. Isomerization process in protonated cyclopropane and cyclobutane ions. *J. Am. Chem. Soc.* **92**, 6430–6440 (1970).
19. Vogel, P. *Carbocation Chemistry* 331–335 (Elsevier, Amsterdam, 1985).
20. Koch, W., Liu, B. & Schleyer, P. V. R. Definitive characterization of the C₃H₇⁺ potential energy surface. *J. Am. Chem. Soc.* **111**, 3479–3480 (1989).
21. Frisch, M. J. *et al.* Gaussian98, Revision A.6 (Gaussian, Inc., Pittsburgh, Philadelphia, 1998).
22. Pedley, J. B., Naylor, R. D. & Kirby, S. P. *Thermochemical Data of Organic Compounds* 215 (Chapman and Hall, London, 1986).
23. Chong, S. L. & Franklin, J. L. Heats of formation of protonated cyclopropane, methylcyclopropane, and ethane. *J. Am. Chem. Soc.* **94**, 6347–6351 (1972).
24. Chase, M. W. Jr *et al.* JANAF Thermochemical Tables Third Edition. *J. Phys. Chem. Ref. Data* **14** (suppl. 1) 1332 (1985).

Acknowledgements

This work was supported by CRI, the Ministry of Science and Technology, Republic of Korea. S.T.P. thanks the Ministry of Education, Republic of Korea, for support through the Brain Korea 21 program.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.S.K. (e-mail: myungsoo@plaza.snu.ac.kr).

Scaling effects in caudal fin propulsion and the speed of ichthyosaurs

Ryosuke Motani

Royal Ontario Museum, Department of Palaeobiology, 100 Queen's Park, Toronto, Ontario M5S 2C6, Canada

Four unrelated groups of large cruising vertebrates (tunas, whales, lamnid sharks and parvipelvian ichthyosaurs) evolved tuna-shaped (thunniform) body plans^{1,2}. Stringent physical constraints, imposed by the surrounding fluids, are probably responsible for this example of evolutionary convergence^{1,2}. Here I present a mathematical model of swimming kinematics and fluid mechanics that specifies and quantifies such constraints, and test the model with empirical data. The test shows quantitatively that morphology, kinematics, and physiology indeed covary tightly in large cruisers. The model enables calculations of optimal cruising speed from external measurements, and also predicts that wide caudal fin spans, typical of thunniform swimmers, are necessary for large cruisers. This finding is contrary to a popular yet rather teleological view that thunniform tails were selected for their high aspect ratios that increased propulsive efficiency^{1,2}. I also show by calculation that *Stenopterygius*, a Jurassic ichthyosaur, probably had optimal cruising speeds and basal metabolic rates similar to living tunas.

Aquatic vertebrates with demarcated caudal fins (see Fig. 1a) produce thrust by forming a reverse von Kármán vortex street in their wakes^{3,4}. The trailing edge of the caudal fin plays an important role in this process by determining the size of the wake^{4,5}. When swimming at a constant speed, the time-averaged force output of the trailing edge that is used as thrust (F_{te}) may be expressed by the equation

$$F_{te} = C_{tte} 0.5 \rho S_{te} U_{te}^2$$

where C_{tte} is the swept-area thrust coefficient of the trailing edge, S_{te} is the area swept by the trailing edge, U_{te} is the time-averaged magnitude of the speed of the trailing edge, and ρ is the density of

the surrounding water. Then, P , the time-averaged power output of the trailing edge used for thrust generation, is described as $P_{te} = F_{te} U_{te}$.

The aquatic vertebrates being discussed share similar general body shapes, despite variation, and the use of the caudal fins as oscillating foils. Therefore, it is expected that the power outputs from their caudal fin trailing edges are approximately similar at a given Reynolds number, Re , of the body when there is no acceleration or deceleration: $Re = UL/\nu$, where U is the speed of the animal, L is the fork length (snout tip to caudal fin fork), and ν is the kinematic viscosity of sea water. This correlation is expected to be exponential, like many others in fluid mechanics, and may be given as:

$$P_{te} = C_{tte} 0.5 \rho S_{te} U_{te}^3 = m Re^n$$

where m and n are constants. C_{tte} is expected to be Re -dependent (that is, $C_{tte} = q Re^r$, where q and r are constants), as are many constants in fluid dynamics, so the equation can be simplified to $S_{te} U_{te}^3 = \nu Re^w$, where ν and w are constants. When the water temperature is approximately constant at 20 °C, this can be further simplified by removing the kinematic viscosity:

$$S_{te} U_{te}^3 = b (UL)^a \quad (1)$$

where a and b are constants.

This hypothesis can be tested by examining whether the correlation given in equation (1) is supported by empirical data. I compiled published kinematic data for vertebrates with demarcated caudal fins^{6–13} during steady swimming (see Methods). Figure 1b shows that there is a high correlation between UL and $S_{te} U_{te}^3$ ($r^2 = 0.983$; $P < 0.001$). This is partly because U_{te} is not entirely independent of U , but neither U_{te} nor S_{te} independently shows correlation with UL as high as $S_{te} U_{te}^3$ (Fig. 1c). Therefore, the dependence of U_{te} on U cannot explain the observed high correlation between UL and $S_{te} U_{te}^3$, which indicates that it is the particular combination $S_{te} U_{te}^3$ that is significant. Thus the hypothesis described in equation (1) passes the empirical test.

Equation (1) describes a simplified model of scaling effects in the power output from the caudal fin trailing edge during steady swimming. The equation is complicated because propulsion by oscillating foils involves a dynamic equilibrium among interacting components⁵. This equation has implications for the scaling of optimal cruising speeds (U_{opt} , the speed at which the energy

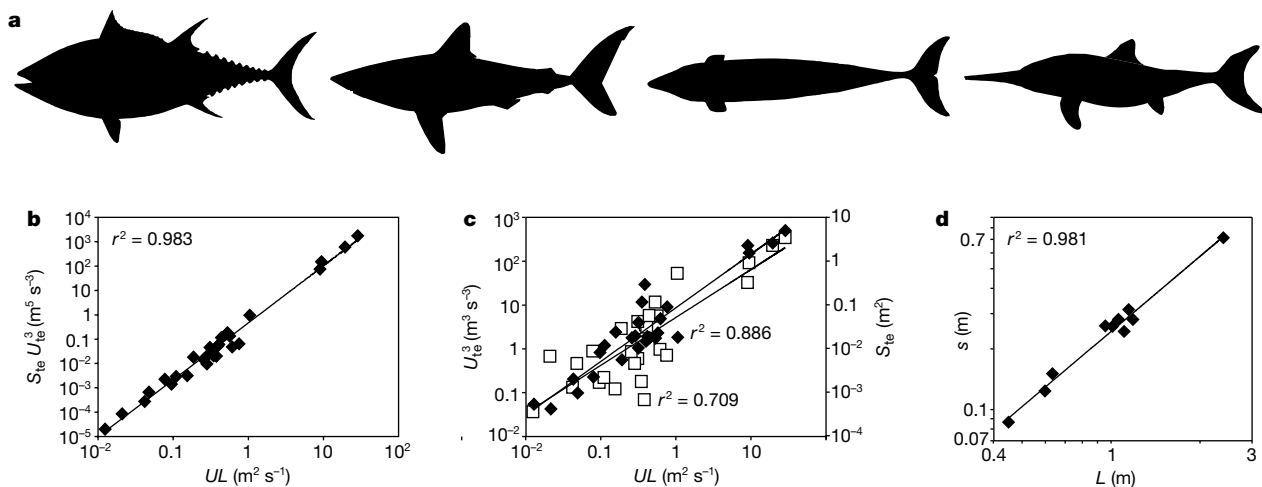


Figure 1 Observed correlations. **a**, Correlation among body profile and cruising habit in large pelagics. From left to right: bluefin tuna (*Thunnus thynnus*), porbeagle (*Lamna nasus*), Heaviside's dolphin (*Cephalorhynchus heavisidii*) and the ichthyosaur *Stenopterygius quadriscissus*. **b** and **c**, Scaling relationships between UL and $S_{te} U_{te}^3$ (**b**) and U_{te}^3 (white squares, **c**) and S_{te} (black diamonds, **c**). I note that the particular

combination of S_{te} and U_{te} in **b** results in a high correlation, despite much lower correlations seen in its components (**c**). **d**, Fluke span, s , versus fork length, L , of *Stenopterygius* ($s = 0.247 \times L^{1.24}$). U , speed of animal; S_{te} , area swept by trailing edge of caudal fin; U_{te} , time-averaged speed of trailing edge.

Table 1 Comparison of optimal swimming speeds

Taxa	Fluke length (m)	Published U_{opt} (m s ⁻¹)	Present estimates of U_{opt} (m s ⁻¹)		Reference
			Equation (3)	Equation (4)	
<i>Thunnus albacares</i>	0.51	1.0	1.0 (0.89–1.2)	1.3 (1.2–1.5)	20
<i>Tursiops truncatus</i>	2.43*	2.1/2.5	2.1 (1.9–2.4)	2.7 (2.4–3.2)	26, 27
<i>Balaenoptera acutorostrata</i>	7.25*	3.25/3.49	3.4† (2.9–4.1)	4.4† (3.7–5.4)	28, 29
<i>B. physalus</i>	14.5	6.0–8.0‡	6.2 (5.0–7.8)	8.0 (6.4–10)	15
<i>Makaira mazara</i>	1.8§	1.0	1.0¶ (0.89–1.1)	1.3¶ (1.2–1.5)	16

Confidence intervals ($P = 0.05$) are given in parentheses for each estimated speed.

* Median of the animals examined.

† Fluke span was isometrically scaled from an individual with a fork length of 6.14 m (ref. 30).

‡ Based on fluke hydrodynamics rather than energetics.

§ Length from lower jaw to caudal fin fork (called body length in billfishes²⁹).

|| Fastest speed sustained for half an hour, not the optimal speed.

¶ Caudal fin span based on the outline of an individual with body length 1.86 m (ref. 23).

consumption needed to move the body a unit length is minimal). High propulsive efficiencies of oscillating foils are achieved at Strouhal numbers (St , usually defined as Af/U for oscillating foils³, where A is the double heaving amplitude of the caudal fin and f is the tailbeat frequency) between 0.25 and 0.4 (ref. 5), whereas the apparent minimization of drag occurs at St numbers of between 0.12 and 0.35 (ref. 14). Vertebrates with demarcated caudal fins are expected to use Strouhal numbers between 0.25 and 0.35 while cruising efficiently, and published data support this hypothesis³. Therefore, at U_{opt} it is likely that:

$$0.35 \geq St_{U_{opt}} \geq 0.25 \quad (2)$$

Solving equations (1) and (2) simultaneously enables two unknown values, U_{opt} and $f_{U_{opt}}$ (the tailbeat frequency at U_{opt}), to be calculated (see Methods). I tested this model with published optimal swimming speeds of vertebrates with demarcated caudal fins. U_{opt} values are available from physiological studies of three species of thunniform swimmers, ranging in Re from about 5×10^5 to 1×10^7 , and the present method approximated these published values well (Table 1). The method was also successful in approximating the U_{opt} value of a fin whale calculated for a linear model of fluke hydrodynamics¹⁵, with Re of about 10^8 (Table 1), although this is an extrapolation given the range of the data used in Fig. 1b. In addition, the method suggests a low U_{opt} value of about 1.0 m s^{-1} for a Pacific blue marlin, which accords well with observed swimming speeds¹⁶ (Table 1). These results show that the generalized predictions given by the present method are still useful in individual cases.

Given the high predictive value of the proposed method in estimating U_{opt} of extant vertebrates, it seems possible to apply it to extinct forms. Several specimens of the ichthyosaur *Stenopterygius* from the Toarcian (Lower Jurassic) of Germany preserve approximate body outlines¹². Artificial modifications of these outlines have been suspected by some¹⁷, but the problem is probably not universal². At least the fluke spans seem to be more

reliable: it is unlikely that the observed positive allometry (Fig. 1d) is artificial. U_{opt} values were estimated for ten individuals of *Stenopterygius*, ranging from 0.450 to 2.40 m in fork length (Table 2). The estimated speeds are slightly slower than those for living cetaceans of similar size (Table 2).

Swimming speed has previously been estimated for ichthyosaurs using an energetic model that required metabolic rates to be assumed a priori^{18,19}. The metabolic rates of ichthyosaurs, of course, are unknown, so estimates were made for three basal metabolic rate categories known in extant vertebrates that might be appropriate for ichthyosaurs. These are the average ectothermic

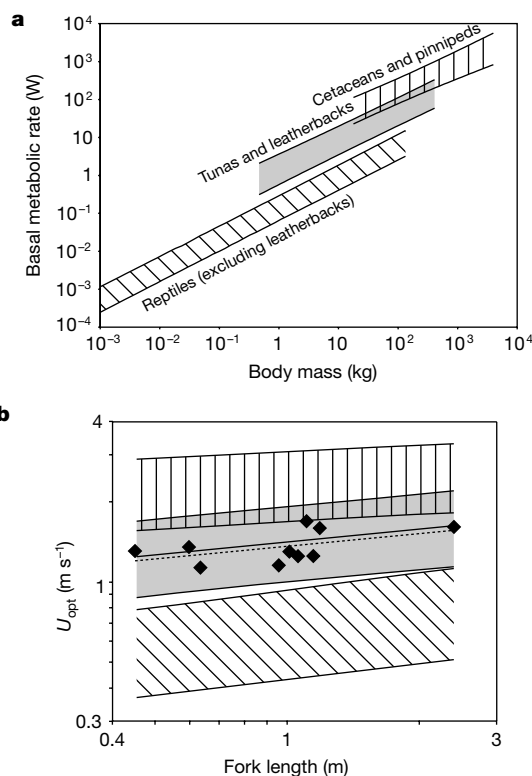


Figure 2 Metabolic rates and speed. **a**, Three basal metabolic rate models of extant amniotes given as prediction belts ($P = 0.05$; ambient temperature about 20°C). The three are reptiles excluding leatherback turtles (oblique hatching), tunas and leatherback turtles (grey), and cetaceans and pinnipeds (vertical hatching). Data sources as in ref. 19. **b**, Present speed estimates for *Stenopterygius* compared with published energetic estimates¹⁹. Black diamonds representing the present estimates (with solid regression line) are superimposed on prediction belts ($P = 0.05$) of speed estimates¹⁹ assuming the three models of metabolic rates given in **a**. I note that the present estimates overlap only the grey area.

Table 2 Estimated optimal swimming speeds for *Stenopterygius*

Specimen number	Fluke length (m)	Estimated U_{opt} (m s ⁻¹)
SMNS Small	0.450	1.3 (1.2–1.5)
SMNS 54818	0.598	1.4 (1.2–1.6)
PMU R435	0.634	1.2 (1.0–1.3)
MHH 9	0.957	1.2 (1.0–1.3)
SMNS 56631	1.01	1.3 (1.2–1.5)
GPIT Q07	1.06	1.3 (1.1–1.4)
GPIT Re1297/1	1.11	1.7 (1.5–1.9)
BSPM 2	1.15	1.3 (1.1–1.4)
PMU R158	1.19	1.6 (1.4–1.8)
SMNS 16811	2.40	1.6 (1.5–1.9)

All measurements are taken from unpublished photographs. Confidence intervals ($P = 0.05$) are given in parentheses for each estimated speed. The double heaving amplitude, A , was assumed to be 0.2L when no data were available. Institutional abbreviations: BSPM, Bayerische Staatssammlung für Paläontologie und historische Geologie, Munich; GPIT, Geologisch-Paläontologisches Institut Tübingen; MHH, Museum Hauff, Holzmaden; PMU, Paleontologiska Museet, Uppsala Universitet; SMNS, Staatliches Museum für Naturkunde, Stuttgart.

level seen in reptiles, the raised ectothermic level in tunas and the high marine mammalian level in cetaceans and pinnipeds (Fig. 2a)¹⁹. The present results closely match the energetic estimates that assumed the basal metabolic rates of tunas for *Stenopterygius* (Fig. 2b). There is no statistically significant difference between the regression lines from the two sets of estimates ($P < 0.01$ for both slope and intercept). The two methods are based on different assumptions and measurements, and therefore are independent of each other. Given that both methods successfully approximate the optimal speeds of living vertebrates, the simplest interpretation of Fig. 2b is that *Stenopterygius* had a 'raised ectothermic' level of basal metabolic rates. This indication is unsurprising because many extant thunniform swimmers have raised basal metabolic rates^{2,20}.

Equations (1) and (2) predict that a larger caudal fin span for a given body length, as in thunniform swimmers, leads to lower U_{opt} and $f_{U_{opt}}$ values. This may seem contradictory to the generally accepted view that the thunniform body plan enables fast cruising^{1,2}. However, it does not mean that thunniform swimmers are slow. The optimal contraction frequency of muscles becomes lower with increasing body size^{21,22}. It is therefore reasonable for large cruisers to have lower $f_{U_{opt}}$ values that match their muscle physiology, which can be achieved by a large caudal fin span. This may be one of the reasons why the thunniform body plan is common in large cruisers². Also, the observed positive allometry of caudal fin span with body size in *Stenopterygius* (Fig. 1d) and some extant non-mammalian thunniform swimmers (for example, refs 23 and 24) may be related to the hydrodynamic and physiological scaling effects discussed above.

Placing vertebrate swimming kinematics in the perspective of scaling effects allows large-scale patterns to emerge, providing frameworks for integrated evolutionary studies. This approach complements the established use of relative speeds (body length per second) and other normalized speeds²⁵ in kinematic studies of swimming vertebrates. □

Methods

Kinematic data

Kinematic data were taken from the literature^{6–13}, by digitizing published graphs when necessary. When total lengths were given in the literature, they were converted to fork lengths with regression equations posted at FishBase (<http://www.fishbase.org>) for respective species. Fork length was preferred to total length because of its prevalence in cetacean literature (which seldom lists conversion factors between total and fork lengths). When the span of the caudal fin (or fluke) was not given in the literature, it was calculated by obtaining the ratio of the span versus fork length from illustrations posted at FishBase.

The average for each size class of each species was plotted in Fig. 1b. Twelve species were used: *Delphinapterus leucas*⁶, *Orcinus orca*⁶, *Pseudorca crassidens*⁶, *Trachurus symmetricus*⁷, *Pollachius virens*⁸, *Scomber scombrus*⁸, *Isurus oxyrinchus*⁹, *Triakis semifasciata*⁹, *Tursiops truncatus*¹⁰, *Thunnus albacares*¹¹, *Scomber japonicus*¹² and *Onchorhynchus mykiss*¹³. Data points with Strouhal numbers higher than 0.4 or lower than 0.12 were omitted. Ambient temperature varied among the studies assessed, but was mostly about 20 °C, at which the kinematic viscosity of seawater is about 1.0×10^{-6} (ref. 1). I therefore assumed a constant temperature of 20 °C to enable compilation of data. Data for the twelve species ($n = 27$) gave a least squares regression equation:

$$S_{ic} U_{ic}^3 = 0.468(UL)^{2.34} \quad (r^2 = 0.983; P < 0.001) \quad (3)$$

Optimal speed

U_{ic} and S_{ic} in equations (1) and (3) are given as:

$$U_{ic} = \{(2Af)^2 + U^2\}^{0.5}$$

$$S_{ic} = s \int_0^{L_s} [1 + \{A\pi/L_s \times \cos(2\pi x/L_s)\}^2]^{0.5} dx$$

where s is the span of the trailing edge and L_s the stride length of swimming (U/f). Therefore, equations (2) and (3) contain five variables: A , s , L_s , U_{opt} and $f_{U_{opt}}$. Accordingly, it is possible to obtain U_{opt} and $f_{U_{opt}}$ by substituting empirical values of L , s and A into the simultaneous equations.

The use of equation (3) led to slight overestimations of published U_{opt} values (Table 1). This is probably because the kinematic data used do not represent the true distribution of the variables in nature, although they approximate it reasonably well. When fitting

equation (1) to give the least squares error with the published U_{opt} , the equation becomes:

$$S_{ic} U_{ic}^3 = 0.395(UL)^{2.34} \quad (4)$$

Equation (4) is indistinguishable from equation (3) when plotted on Fig. 1b, and was used instead of the latter for estimating U_{opt} . Confidence intervals, CI, of this equation were calculated by fitting raw kinematic data to equation (4).

$$S_{ic} U_{ic}^3 = 0.395(UL)^{2.34} \times 10^{\pm CI}$$

$$CI = 0.496 \times \{3.60 \times 10^{-3} + 3.73 \times 10^{-3} [\log(UL) - 0.245]^2\}^{0.5} \quad (P = 0.05) \quad (5)$$

Confidence ranges for U_{opt} were calculated on the basis of equation (5).

Ichthyosaurs

Length measurements for *Stenopterygius* were taken from photographs of fossil specimens with body outlines, using ImageTool 2.0 (<http://ddsdx.uthscsa.edu/dig/>). The double heaving amplitude, A , cannot be obtained from a fossil, but A is about 20% of the fork length in odontocetes, with a very limited variation⁶, and similar values have been reported for teleosts. Therefore, for vertebrates cruising at U_{opt} , it seems reasonable to assume $A = 0.2L$ when no data are available.

Received 17 July; accepted 19 November 2001.

- Vogel, S. *Life in Moving Fluids. The Physical Biology of Flow* 2nd edn (Princeton Univ. Press, Princeton, 1994).
- McGowan, C. *Dinosaurs, Spitfires, and Sea Dragons* (Harvard Univ. Press, Cambridge/London, 1991).
- Triantafyllou, G. S., Triantafyllou, M. S. & Grosenbaugh, M. A. Optimal thrust development in oscillating foils with application to fish propulsion. *J. Fluids Struct.* **7**, 205–224 (1993).
- Wolfgang, M. J., Anderson, J. M., Grosenbaugh, M. A., Yue, D. K. P. & Triantafyllou, M. S. Near-body flow dynamics in swimming fish. *J. Exp. Biol.* **202**, 2303–2327 (1999).
- Anderson, J. M., Streitlien, K., Barrett, D. S. & Triantafyllou, M. S. Oscillating foils of high propulsive efficiency. *J. Fluid Mech.* **360**, 41–72 (1998).
- Fish, F. E. Comparative kinematics and hydrodynamics of odontocete cetaceans: morphological and ecological correlates with swimming performance. *J. Exp. Biol.* **201**, 2867–2877 (1998).
- Hunter, J. R. & Zweifel, J. R. Swimming speed, tail beat frequency, tail beat amplitude, and size in jack mackerel, *Trachurus symmetricus*, and other fishes. *Fish. Bull.* **69**, 253–266 (1971).
- Videler, J. J. & Hess, F. Fast continuous swimming of two pelagic predators, saithe (*Pollachius virens*) and mackerel (*Scomber scombrus*): a kinematic analysis. *J. Exp. Biol.* **109**, 209–228 (1984).
- Graham, J. B., Dewar, H., Lai, N. C., Lowell, W. R. & Arce, S. M. Aspects of shark swimming performance determined using a large water tunnel. *J. Exp. Biol.* **151**, 175–192 (1990).
- Fish, F. E. Power output and propulsive efficiency of swimming bottlenose dolphins (*Tursiops truncatus*). *J. Exp. Biol.* **185**, 179–193 (1993).
- Dewar, H. & Graham, J. B. Studies of tropical tuna swimming performance in a large water tunnel. III. Kinematics. *J. Exp. Biol.* **192**, 45–59 (1994).
- Gibb, A. C., Dickson, K. A. & Lauder, G. V. Tail kinematics of the chub mackerel *Scomber japonicus*, testing the homocercal tail model of fish propulsion. *J. Exp. Biol.* **202**, 2433–2447 (1999).
- Webb, P. W. & Kosteki, P. T. The effect of size and swimming speed on locomotor kinematics of rainbow trout. *J. Exp. Biol.* **109**, 77–95 (1984).
- Barrett, D. S., Triantafyllou, M. S., Yue, D. K. P., Grosenbaugh, M. A. & Wolfgang, M. J. Drag reduction in fish-like locomotion. *J. Fluid Mech.* **392**, 183–212 (1999).
- Bose, N. & Lien, J. Propulsion of a fin whale (*Balaenoptera physalus*): why the fin whale is a fast swimmer. *Proc. R. Soc. Lond. B* **237**, 175–200 (1989).
- Block, B. A., Booth, D. J. & Carey, F. G. Direct measurement of swimming speeds and depth of blue marlin. *J. Exp. Biol.* **166**, 267–284 (1992).
- Martill, D. M. Prokaryote mats replacing soft tissues in Mesozoic marine reptiles. *Mod. Geol.* **11**, 265–269 (1987).
- Massare, J. A. Swimming capabilities of Mesozoic marine reptiles: implications for method of predation. *Paleobiology* **14**, 187–205 (1988).
- Motani, R. Swimming speed estimation of extinct marine reptiles I: Energetic approach revisited. *Paleobiology* (submitted).
- Dewar, H. & Graham, J. B. Studies of tropical tuna swimming performance in a large watertunnel. I. Energetics. *J. Exp. Biol.* **192**, 13–31 (1994).
- Altringham, J. D. & Johnston, I. A. Scaling effects on muscle function: power output of isolated muscle fibres performing oscillatory work. *J. Exp. Biol.* **151**, 453–467 (1990).
- Altringham, J. D. & Young, I. S. Power output and the frequency of oscillatory work in mammalian diaphragm muscle: the effects of animal size. *J. Exp. Biol.* **157**, 381–389 (1991).
- Nakamura, I. FAO species catalogue. Vol. 5. Billfishes of the world. An annotated and illustrated catalogue of marlins, sailfishes, spearfishes and swordfishes known to date. *FAO Fish. Synop.* **125**, 1–65 (1985).
- Arata, G. F. A contribution to the life history of the swordfish, *Xiphias gladius* Linnaeus, from the South Atlantic coast of the United States and the Gulf of Mexico. *Bull. Mar. Sci. Gulf Caribb.* **4**, 183–243 (1954).
- Drucker, E. G. The use of gait transition speed in comparative studies of fish locomotion. *Am. Zool.* **36**, 555–566 (1996).
- Williams, T. M., Friedl, W. A. & Haun, J. E. The physiology of bottlenose dolphins (*Tursiops truncatus*): heart rate, metabolic rate and plasma lactate concentration during exercise. *J. Exp. Biol.* **179**, 31–46 (1993).
- Yazdi, P., Kilian, A. & Culik, B. M. Energy expenditure of swimming bottlenose dolphins (*Tursiops truncatus*). *Mar. Biol.* **134**, 601–607 (1999).
- Blix, A. S. & Folkow, L. P. Daily energy expenditure in free living minke whales. *Acta Physiol. Scand.* **153**, 61–66 (1995).
- Hind, A. T. & Gurney, W. S. C. The metabolic cost of swimming in marine homeotherms. *J. Exp. Biol.* **200**, 531–542 (1997).
- Bose, N., Lien, J. & Ahia, J. Measurements of the bodies and flukes of several cetacean species. *Proc. R. Soc. Lond. B* **242**, 163–174 (1990).

Acknowledgements

I thank F. E. Fish, G. V. Lauder, C. McGowan, K. Padian, H.-D. Sues, and P. W. Webb for their comments and constructive criticisms on earlier versions of the manuscript. This study was supported by a Natural Sciences and Engineering Research Council grant to C. McGowan.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.M. (e-mail: ryo.motani@utoronto.ca).

A hydrogen-based subsurface microbial community dominated by methanogens

Francis H. Chapelle*, Kathleen O'Neill†, Paul M. Bradley*, Barbara A. Methé†, Stacy A. Ciufo†, LeRoy L. Knobel‡ & Derek R. Lovley†

* US Geological Survey, Columbia, South Carolina 29210, USA

† Department of Microbiology, University of Massachusetts, Amherst, Massachusetts 01003, USA

‡ US Geological Survey, Idaho Falls, Idaho 83402, USA

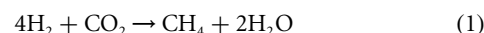
The search for extraterrestrial life may be facilitated if ecosystems can be found on Earth that exist under conditions analogous to those present on other planets or moons. It has been proposed, on the basis of geochemical and thermodynamic considerations, that geologically derived hydrogen might support subsurface microbial communities on Mars and Europa in which methanogens form the base of the ecosystem^{1–5}. Here we describe a unique subsurface microbial community in which hydrogen-consuming, methane-producing Archaea far outnumber the Bacteria. More than 90% of the 16S ribosomal DNA sequences recovered from hydrothermal waters circulating through deeply buried igneous

rocks in Idaho are related to hydrogen-using methanogenic microorganisms. Geochemical characterization indicates that geothermal hydrogen, not organic carbon, is the primary energy source for this methanogen-dominated microbial community. These results demonstrate that hydrogen-based methanogenic communities do occur in Earth's subsurface, providing an analogue for possible subsurface microbial ecosystems on other planets.

The lack of water on the surface of extraterrestrial bodies within our Solar System suggests that, if extraterrestrial life exists in the Solar System, it will be found in the subsurface, where liquid water may be present^{1–7}. These extraterrestrial subsurface habitats are not expected to contain significant quantities of organic carbon to support life. However, the interaction of hydrothermal fluids with igneous rocks can produce hydrogen or other reduced compounds that may provide an energy source for microorganisms^{2,4}.

On Earth, microorganisms can gain energy by coupling the oxidation of hydrogen to the reduction of compounds such as oxygen, nitrate, Fe(III), sulphate or carbon dioxide. Molecular oxygen, which is required for the production of significant quantities of nitrate, Fe(III) and sulphate, is available on Earth because of photosynthesis. Although photolysis also produces oxygen on Earth and on extraterrestrial bodies, it is uncertain whether this process delivers oxygen to subsurface environments. This suggests that carbon dioxide is likely to be the most abundant electron acceptor for hydrogen oxidation in the subsurface of extraterrestrial bodies.

It has been suggested^{1,2} that life in the subsurface of Mars and Europa might consist primarily of microorganisms that couple the oxidation of hydrogen with the reduction of carbon dioxide to produce methane according to the reaction



In an ecosystem where they form the base of the food chain, hydrogen-consuming methanogenic microorganisms would be expected to be numerically predominant, although heterotrophic microorganisms may also be present in lesser numbers.

To find a subsurface environment that might serve as a suitable analogue for proposed hydrogen-based microbial ecosystems on Mars and Europa, we searched for a subsurface environment that lacked organic carbon, but that had a source of geologically produced hydrogen. These conditions occur at Lidy Hot Springs, Idaho, which is located at the southern extent of the Beaverhead Mountains. The underlying rocks are of volcanic origin and are devoid of allochthonous organic carbon⁸, and there are potential sources of geologically produced hydrogen^{9–11}. Moreover, there is a well developed system of geothermal springs that tap deep fault zones^{12,13}. This deep (200 m below land surface) hydrothermal

Table 1 Chemical composition of groundwater from Lidy Hot Springs

Constituent or property	Value
Temperature (°C)	58.5
Microorganisms (cells per ml)	$(2.8 \pm 0.7) \times 10^5$
pH	6.77
Dissolved oxygen (mg l ⁻¹)	<0.05
Iron (mg l ⁻¹)	<0.01
Sulphide (mg l ⁻¹)	0.070
Methane (mg l ⁻¹)	2.0
Hydrogen (nM)	13.0 ± 2.0
Carbon monoxide (nM)	1.3
Dissolved inorganic carbon (mg l ⁻¹)	666
Dissolved organic carbon (mg l ⁻¹)	<0.27
¹⁴ C of dissolved inorganic carbon (per cent of modern value)	<1.20
$\delta^{13}\text{C}$ of dissolved inorganic carbon (per ml)	-2.1 ± 0.2
Formate (μM)	<1.0
Acetate (μM)	<1.0
Propionate (μM)	<1.0
Butyrate (μM)	<1.0
Sulphate (mg l ⁻¹)	131
Nitrate (mg l ⁻¹)	<0.1
Chloride (mg l ⁻¹)	6.4
Calcium (mg l ⁻¹)	91.0
Magnesium (mg l ⁻¹)	21.5
Potassium (mg l ⁻¹)	17.1
Sodium (mg l ⁻¹)	38.3
Ammonium (mg l ⁻¹)	0.56
Alpha (as ²³⁰ Th; g l ⁻¹)	14.4 ± 1.17
Beta (as ⁹⁰ Sr; pCi l ⁻¹)	12.9 ± 1.03
²²² Rn (pCi l ⁻¹)	470 ± 14

Values with error are mean $\pm$ s.d.

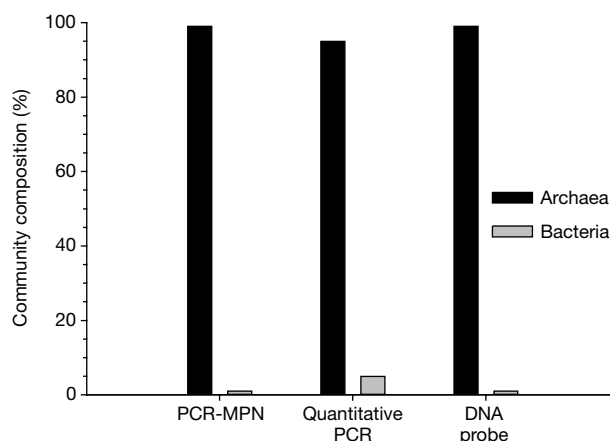


Figure 1 Proportions of Archaea and Bacteria in groundwater from Lidy Hot Springs, as determined by three different molecular techniques.

system was specially instrumented to recover water samples free from surface contamination.

The subsurface water was hot (58.5°C) and anoxic, and contained negligible concentrations of dissolved organic carbon and short-chain fatty acids (Table 1). Even acetate, the predominant organic acid found in anaerobic environments in which organic matter serves as the primary energy source, was undetectable (Table 1). The low ^{14}C and slightly negative $\delta^{13}\text{C}$ content of dissolved inorganic carbon (Table 1) indicated an ancient (>15,000 years), deep source for the water, which is consistent with the mantle-derived 'hot spot' origin of volcanism in this part of Idaho¹². This further indicated that the water did not contain recent surface-derived organic carbon capable of serving as an energy source for microorganisms.

In contrast to the lack of organic electron donors for microbial metabolism, the water did contain molecular hydrogen (Table 1) at concentrations previously shown to support the activity of hydrogen-oxidizing methanogenic microorganisms in a wide diversity of sedimentary environments¹⁴. Hydrogen is commonly associated with hydrothermal waters in volcanic terrains^{9,10} and is produced by active tectonism^{10,11}.

Microscopic cell counts indicated that microorganisms were living at depth in this hydrothermal system (Table 1). To initially characterize this microbial community, microorganisms from the ground water were collected on filters¹⁵, and the number of Bacteria and Archaea were estimated by polymerase chain reaction accom-

panied by a most probable number method (PCR-MPN)¹⁶. Estimating the relative number of Bacteria and Archaea in this manner might be expected to overestimate the importance of Bacteria, because Bacteria are more likely than Archaea to have multiple copies of 16S rRNA genes¹⁷. However, the numbers of Bacteria and Archaea in the water were estimated to be $1,230 \pm 26$ and $227,000 \pm 428$ per ml (mean $\pm$ s.d.) of water respectively. This indicated that Archaea comprised as much as 99% of the microbial community (Fig. 1). Additional analysis of the relative proportions of Archaea and Bacteria with the GeneAmp sequence detection system for quantitative PCR confirmed that the microbial population consisted of more than 95% Archaea (Fig. 1). To evaluate the composition of the microbial community with a technique that did not involve potential biases associated with PCR, samples were also analysed with four radiolabelled probes designed to hybridize with 16S rDNA of either all microorganisms, Archaea, Bacteria or the bacterial order Planctomycetales. When compared with the response to a universal probe designed to hybridize to all microorganisms, about 99% of the 16S rDNA could be attributed to Archaea and 1% hybridized to the probe for Bacteria (Fig. 1). A probe specific for Planctomycetales gave no response.

To learn more about the Archaea present in Lidy Hot Springs, 16S rDNA recovered from the waters was cloned and sequenced^{18,19}. The ten clones that were initially sequenced all belonged to two similar sequences (IUA5 and IUA6), which are most closely related to known methanogenic microorganisms (Fig. 2). To further evaluate

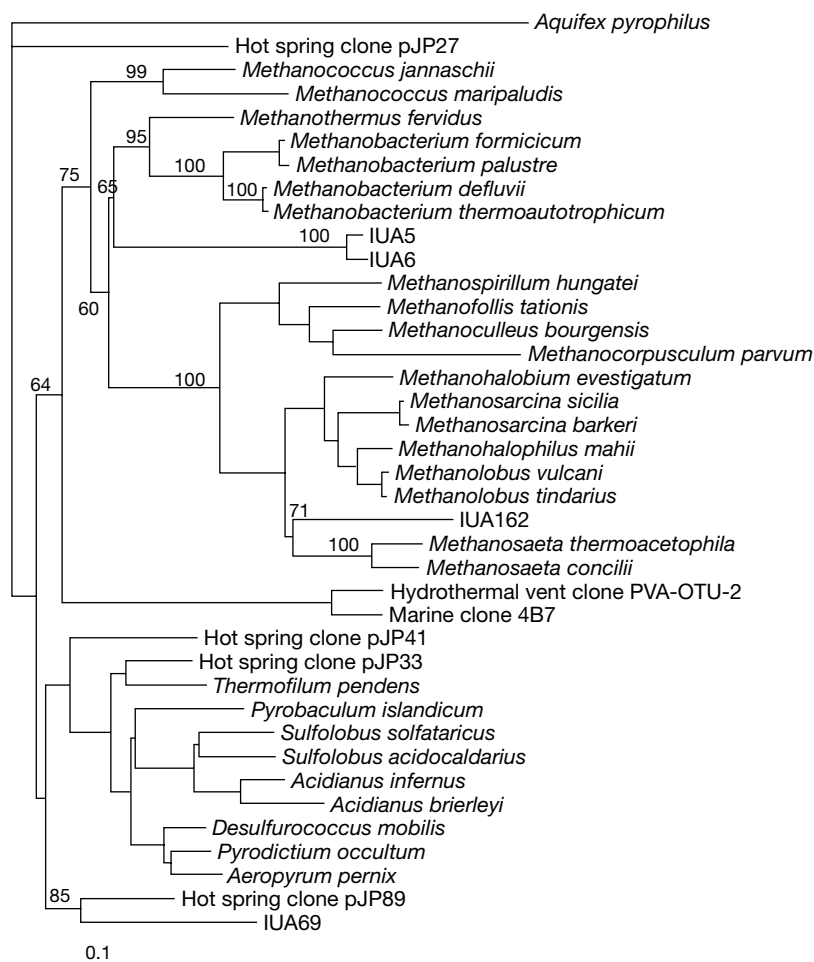


Figure 2 Phylogenetic analysis of archaeal sequences from Lidy Hot Springs. Lidy sequences are designated IUA. The tree was constructed by Jukes–Cantor distance analysis; bootstrap values are shown for key branches.

the predominance of methanogens, 55 additional clones were screened by 16S restriction fragment length polymorphism (RFLP) analysis²⁰. 44 of the clones had RFLP patterns identical to IUA5 and IUA6. An additional 8 of the remaining 11 clones had an identical RFLP pattern, designated IUA162, that although different from IUA5 and IUA6 was also related to known methanogens (Fig. 1). The remaining three sequences, designated IUA69, were most closely related to a sequence recovered from a hot spring in Yellowstone National Park²¹, California, the physiology of which is unknown. Thus, no fewer than 62 of the 65 archaeal sequences recovered from deep circulating hydrothermal waters underlying Lidy Hot Springs seemed to be from methanogenic microorganisms. To our knowledge, such a predominance of methanogens is unknown in subsurface microbial ecosystems.

It has been suggested that hydrogen produced from water reacting with basalt was the main energy source supporting the microbial community in methanogenic portions of the Columbia River basalt aquifer²². However, analysis of 16S rDNA sequences in the groundwater from the site, with the same probes used for analysing the Lidy Hot Springs, demonstrated that less than 3% of the microbial community was composed of methanogenic microorganisms²³, a proportion similar to that found in anaerobic environments in which organic matter is the primary source of energy supporting microbial growth. A likely explanation for this is that little hydrogen is produced from water–basalt interactions under ambient conditions in the Columbia River basalt aquifer, and that organic matter serves as the primary source of energy in this environment²⁴.

The methanogen-dominated microbial community present at Lidy Hot Springs is unlike any previously described on Earth, and the predominance of hydrogen-oxidizing methanogens in this hydrogen-containing subsurface environment, poor in organic carbon, is consistent with geochemical scenarios proposed for microbial communities that may inhabit the subsurface of Mars and Europa^{1–6}. Therefore, terrestrial ecosystems such as Lidy Hot Springs provides a useful analogue for the kinds of hydrogen-based, subsurface microbial communities that may exist elsewhere in the Solar System. □

Methods

Groundwater samples

The well instrumentation was designed and built by C. E. Wilson, the owner of Lidy Hot Springs, as a means to capture and regulate spring flow. At the request of the US Geological Survey, C. E. Wilson modified the instrumentation to enable representative water samples to be obtained directly from deep fracture zones. A hole was drilled through the overburden (~25 m thick), and the overburden was sealed off with steel well casing cemented to the top of the underlying igneous rocks. A drill-rod assembly was equipped with a valve system for restraining the pressure of the hydrothermal waters, and an open bore hole was drilled about 200 m into the underlying igneous rocks. About 100 m of polyethylene casing was then inserted into the bore hole, extending about 75 m below the steel casing into the fractured igneous rocks. Because there is a constant upward flow of groundwater through the plastic casing (upward pressure at the top of the casing is ~275 kPa, or ~40 p.s.i.), and because this water flows continuously from the deep fractures, this configuration makes it possible to obtain water samples suitable for detailed geochemical and microbiological characterization of the deep hydrothermal ecosystem.

Geochemical analyses

Anions, cations and volatile fatty acids were analysed by the US Geological Survey using ion chromatography (anions, volatile fatty acids) and inductively coupled plasma spectroscopy (cations). Dissolved inorganic carbon (DIC) and methane were determined by headspace gas chromatographic analysis of water samples injected into sealed serum bottles. DIC was measured after acidifying the sample. Dissolved hydrogen was measured by the bubble-strip method²⁵. Water samples were screened for the presence of volatile and semi-volatile organic compounds with gas chromatography–mass spectrometry, but no compounds were detected other than internal standards. Microorganisms in the groundwater were enumerated on black filters with epifluorescent microscopy²⁶.

Microbiological characterization

For molecular studies, groundwater was filtered as described previously and DNA extracted¹⁵. For enumeration of Archaea with PCR-MPN¹⁶, 16S rDNA was initially amplified with the primers 25F and 1392R followed by a second round of amplification

with 344F and 915R. To enumerate Bacteria, the initial amplification was with 8F and 1392R followed by amplification with 338F and 907R. This enumeration technique was verified with standards of archaeal and bacterial DNA. Quantitative PCR was performed with the GeneAmp 5700 sequence-detection system (ABI). DNA was amplified with the Bacteria-specific primer 338F or the Archaea-specific primer 344F and the universal primer 907R. Incorporation of the fluorescent dye SybrGreen into double-stranded DNA provided a fluorescent signal, which was detected with the GeneAmp 5700. Relative proportions of archaeal or bacterial targets were determined by interpolation from a standard curve generated with standard additions of archaeal or bacterial DNA. Oligonucleotide probing was carried out as previously described¹⁹. Samples were hybridized with either the universal probe UNIV-1392, archaeal probe Arch-915, bacterial probe EUB-338 or *Planctomycetes* probe PLA886.

Phylogenetic analyses

For phylogenetic analysis, DNA from the hot spring was recovered and amplified as described above, cloned, and sequenced by standard procedures¹⁵. BLAST search (<http://www.ncbi.nlm.nih.gov/BLAST>) was used to identify and obtain similar sequences from GenBank (<http://www.ncbi.nlm.nih.gov/Genbank>)¹⁸. Distance matrix, maximum likelihood and bootstrap analyses were performed using 508 homologous bases for the archaeal clones. For RFLP analysis, 55 clones were amplified with M13 forward and reverse primers. The resulting PCR product was digested overnight with the restriction enzyme *MspI*. Digests were run on a 3% Metaphor gel (FMC Products) and restriction patterns were visualized with ethidium bromide staining.

Received 23 October; accepted 7 November 2001.

1. Boston, P. M., Ivanov, M. V. & McKay, C. P. On the possibility of chemosynthetic ecosystems in subsurface habitats on Mars. *Icarus* **95**, 300–308 (1992).
2. McCollom, T. M. Methanogenesis as a potential source of chemical energy for primary biomass production by autotrophic organisms in hydrothermal systems on Europa. *J. Geophys. Res.* **104** (E12), 30729–30742 (1999).
3. Carr, M. H. in *Evolution of Hydrothermal Ecosystems on Earth (and Mars?)* (eds Bock, G. R. & Goode, J. A.) 249–265 (Wiley, Chichester, 1996).
4. McKay, C. P. in *Subsurface Microbiology and Biogeochemistry* (eds Fredrickson, J. K. & Fletcher, M. J. A.) 315–327 (Wiley, New York, 2001).
5. Fisk, M. R. & Giovannoni, S. J. Sources of nutrients and energy for a deep biosphere on Mars. *J. Geophys. Res.* **104** (E12), 11805–11815 (1999).
6. Shock, E. L. in *Evolution of Hydrothermal Ecosystems on Earth (and Mars?)* (eds Bock, G. R. & Goode, J. A.) 40–60 (Wiley, Chichester, 1996).
7. Farmer, J. D. Hydrothermal systems: Doorways to early biosphere evolution. *GSA Today* **10** (7), 1–9 (2000).
8. Embree, G. F., McBroome, L. A. & Soherty, D. J. Preliminary stratigraphic framework of the Pliocene and Miocene rhyolite, Eastern Snake River Plain. *Idaho Bur. Mines Geol. Bull.* **26**, 333–346 (1982).
9. Sugisaki, R. & Sugiura, T. Gas anomalies at three mineral springs and a fumarole before an inland earthquake, Central Japan. *J. Geophys. Res.* **91**, 12296–12304 (1986).
10. Lilley, M. D. & Olson, E. J. in *Abstr. 11th Annu. V. M. Goldschmidt Conf.* Abstract no. 3682, LPI contribution no. 1088 (Lunar and Planetary Institute, Houston, 2001) (CD-ROM).
11. Wakita, H. *et al.* Hydrogen release: New indicator of fault activity. *Science* **210**, 188–190 (1980).
12. Pierce, K. L. & Morgan, L. A. The track of the Yellowstone hot spot: Volcanism, faulting, and uplift. *Geol. Soc. Am. Mem.* no. 179 (1992).
13. Knobel, L. L. *et al.* Chemical constituents in ground water from 39 selected sites with an evaluation of quality assurance data, Idaho National Engineering and Environmental Laboratory and vicinity, Idaho. *US Geol. Surv. Op. File Rep.* no. 99-246 (1999).
14. Lovley, D. R. & Goodwin, S. Hydrogen concentrations as an indicator of the predominant terminal electron accepting reactions in aquatic sediments. *Geochim. Cosmochim. Acta* **52**, 2993–3003 (1988).
15. Orphan, V. J., Taylor, L. T., Hafenbrandl, D. & Delong, E. F. Culture-dependent and culture-independent characterization of microbial assemblages associated with high-temperature petroleum reservoirs. *Appl. Environ. Microbiol.* **66**, 700–711 (2000).
16. Picard, C. *et al.* Detection and enumeration of bacteria in soil by direct DNA extraction and polymerase chain reaction. *Appl. Environ. Microbiol.* **58**, 2717–2722 (1992).
17. Klappenbach, J. A. *et al.* rrndb: the ribosomal RNA operon copy number database. *Nucleic Acids Res.* **29**, 181–184 (2001).
18. Amann, R. L., Ludwig, W. & Schleifer, K. H. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev.* **59**, 143–169 (1995).
19. Massana, R., Murray, A. E., Preston, C. M. & Delong, E. F. Vertical distribution and phylogenetic characterization of marine planktonic Archaea in the Santa Barbara channel. *Appl. Environ. Microbiol.* **63**, 50–56 (1997).
20. Moyer, C. L., Dobbs, F. C. & Karl, D. M. Estimation of diversity and community structure through restriction fragment length polymorphism distribution analysis of bacterial 16S rRNA genes from a microbial mat at an active hydrothermal vent system, Loihi Seamount, Hawaii. *Appl. Environ. Microbiol.* **60**, 871–879 (1994).
21. Barns, S. M. *et al.* Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. *Proc. Natl Acad. Sci. USA* **91**, 1609–1613 (1994).
22. Stevens, T. O. & McKinley, J. P. Lithoautotrophic microbial ecosystems in deep basalt aquifers. *Science* **270**, 450–454 (1995).
23. Fry, N. K. *et al.* Population structure of microbial communities associated with two deep, anaerobic, alkaline aquifers. *Appl. Environ. Microbiol.* **63**, 1498–1504 (1997).
24. Anderson, R. T., Chapelle, F. H. & Lovley, D. R. Evidence against hydrogen-based microbial ecosystems in basalt aquifers. *Science* **281**, 976–977 (1998).
25. Chapelle, F. H., Vroblesky, D. A., Woodward, J. C. & Lovley, D. R. Practical considerations for measuring hydrogen concentrations in groundwater. *Environ. Sci. Tech.* **31**, 2873–2877 (1997).
26. Hobbie, J. E., Daley, R. J. & Jasper, S. Use of Nuclepore filters for counting bacteria by fluorescence microscopy. *Appl. Environ. Microbiol.* **33**, 1225–1228 (1977).

Acknowledgements

This research was supported by the US Geological Survey and a grant from the LexEn programme of the National Science Foundation.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to F.H.C. (e-mail: chapelle@usgs.gov) or D.R.L. (e-mail: dlovley@microbio.umass.edu). GenBank accession numbers for sequences IUA5, IUA6, IUA162 and IUA69 are AF448449, AF448450, AF448 451 and AF448452, respectively.

Contribution of *Distal-less* to quantitative variation in butterfly eyespots

Patrícia Beldade*, Paul M. Brakefield* & Anthony D. Long†

* Institute of Evolutionary and Ecological Sciences, University of Leiden, PO Box 9516, 2300 RA Leiden, The Netherlands

† Department of Ecology and Evolutionary Biology, University of California, Irvine, California 92612, USA

The colour patterns decorating butterfly wings provide ideal material to study the reciprocal interactions between evolution and development. They are visually compelling products of selection, often with a clear adaptive value, and are amenable to a detailed developmental characterization¹. Research on wing-pattern evolution and development has focused on the eyespots of the tropical butterfly *Bicyclus anynana*². There is quantitative variation for several features of eyespot morphology^{3–5} but the actual genes contributing to such variation are unknown. On the other hand, studies of gene expression patterns in wing primordia have implicated different developmental pathways in eyespot formation^{6–11}. To link these two sets of information we need to identify which genes within the implicated pathways contribute to the quantitative variation accessible to natural selection. Here we begin to bridge this gap by demonstrating linkage between DNA polymorphisms in the candidate gene *Distal-less* (*Dll*) and eyespot size in *B. anynana*.

The comparison of gene expression patterns across species has been a common approach in evolutionary developmental biology. This approach can identify steps in developmental pathways that have been altered during evolution, but it fails to identify the actual genetic changes that have occurred¹². Despite the recognized importance of understanding the generation of the variants that can be sorted by natural selection¹³, the genes contributing to standing quantitative variation in morphological traits are largely unknown. Recent work has focused on bristle number in *Drosophila melanogaster*¹⁴. Unfortunately the direct developmental mechanisms through which polymorphisms in candidate genes contribute to variation in bristle number, and the ecological significance of this variation, are difficult to determine. Here we investigate the genetic basis of eyespot size in *B. anynana*, a character of more obvious adaptive significance^{2,15} and whose developmental basis can be dissected using manipulative experiments³.

Studies of gene expression patterns in pre-adult wing primordia have implicated a series of genes in butterfly wing patterning and, in particular, in eyespot formation^{6–11}. Genes including *Dll* and genes of the *hedgehog* pathway have circular regions of expression that correspond to the position of eyespot centres^{8,9} whose organizing properties have been clearly demonstrated^{16,17}. *Dll* is particularly

interesting because its expression patterns appear to reveal different stages in eyespot formation and parallel adult variants of eyespot morphology⁸. These results suggest that *Dll* is involved in regulating the formation and diversity of eyespot patterns in butterfly wings⁸. However, although *Dll* expression patterns implicate the *Dll* protein in eyespot formation, it is unclear how standing variation at this locus contributes to inter-individual variation in eyespot patterns. Here we test whether segregating variation at *Dll* contributes to short-term response to selection on eyespot size.

We used artificial selection to establish lines that differed in the size of the two dorsal forewing eyespots of *B. anynana* (see Methods). After nine generations of selection, the lines show markedly different phenotypes; the 'high' line has large eyespots and the 'low' line, small eyespots (Fig. 1a, b). The rapid and gradual response to selection indicates that there is substantial additive genetic variance for eyespot size. Realized heritabilities of around 0.6–0.7 (Fig. 1a) are comparable to previous estimates for the size of the posterior eyespot³. The selected butterflies also show quantitative differences in *Dll* expression: wing discs of the high line have considerably larger areas of *Dll* expression around the location of the centre of the presumptive eyespots. These differences are already apparent in late final instar larvae (Fig. 1c), and more marked in pupal wing discs (Fig. 1d). The observed ontogeny of *Dll* expression is comparable to that described previously⁸.

The changes in *Dll* expression might be caused by DNA polymorphisms in *Dll* itself or in upstream regulators of *Dll*. To distinguish between the hypotheses of *cis* versus *trans* polymorphisms affecting *Dll* expression, a series of crosses were made using the high and low selection lines (Fig. 2a). The progeny from a backcross between a hybrid butterfly and each of the parental lines have

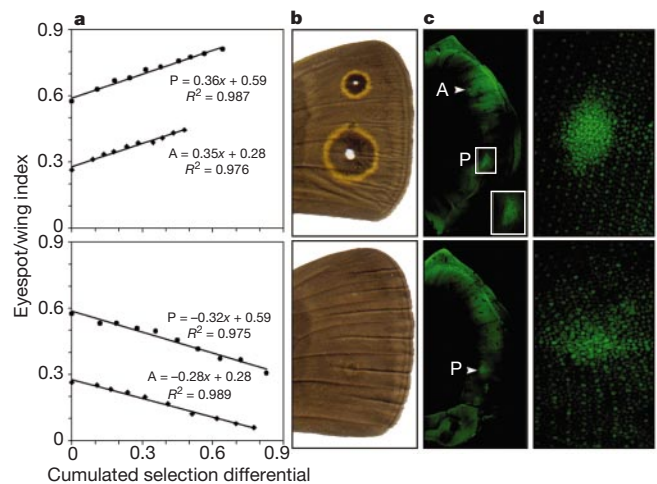


Figure 1 *Bicyclus anynana* selection lines divergent for the size of the dorsal forewing eyespots. **a**, Response to artificial selection for large (high line, top) and small (low line, bottom) eyespots. For each generation, the selection differential is the difference between the mean trait value for all butterflies and that of the selected group²⁹. Realized heritabilities are twice (selection has targeted females only)²⁹ the absolute value of the slopes estimated from least-squares regressions for both eyespots (anterior (A) indicated by diamonds, posterior (P) by circles). **b–d**, Eyespot size phenotypes and *Dll* expression in butterflies from the high and low selection lines. Butterflies from the high line have large adult eyespots (**b**, top). They also show enlarged domains of *Dll* expression in the centre of the putative eyespots both in late larval wing discs (**c**, top; inset with higher magnification of central area of posterior (P) eyespot; arrow points at centre of anterior (A) eyespot) and in pupal wings (anterior eyespot centre shown in **d**, top). Butterflies from the low line have small, often absent, adult eyespots (**b**, bottom) and small areas of *Dll* expression in wing primordia (**c**, **d**, bottom). In the larval disc in **c** (bottom) there is no visible *Dll* expression associated with the anterior eyespot; arrow points to the posterior eyespot centre. In the pupal disc there are small domains of *Dll* expression as shown here for the anterior eyespot (**d**, bottom).

either both *Dll* alleles from the same origin or one allele from each parental line. To distinguish *Dll* alleles from high and low origin we cloned a *B. anynana* *Dll* homologue (GenBank accession number AF404825), which showed 99% amino acid identity with the published *Precis coenia* *Dll*¹⁸. We sequenced 1,250 base pairs (bp) of this gene in individuals from the selection lines (see Methods), and identified 27 segregating sites (corresponding to levels of DNA polymorphism comparable to *Drosophila*¹⁹). We used two of these polymorphisms to distinguish high (H) and low (L) *Dll* alleles and to genotype individuals from the crosses (see Methods). The high backcross produced progeny of genotype HH or HL, and the low backcross produced HL or LL (Fig. 2a). Eyespot sizes of butterflies from distinct genotypic classes were compared to test for an association between *Dll* genotype and that phenotype (see Methods).

We observe that high and low *Dll* alleles segregate with eyespot size in the backcrosses (Fig. 2b). Significant differences were found between the size of the posterior eyespot of HH versus HL backcross females and for both the anterior and posterior eyespots in HL versus LL males (Table 1). Our results show that there are differences in the estimated allelic effects between males and females and between the two target eyespots. Work in *Drosophila* and other organisms has frequently detected this type of sex-specific effect and correlated character-specific effects^{14,20,21}. As the identified *Dll*-linked factor explains only part of the difference between selection lines (Table 1), other genes must have been involved in the response to selection (for ventral eyespot size at least five loci were implicated²²). These genes may include upstream *trans*-regulators of *Dll*.

Segregation of a molecular marker with eyespot size is evidence of linkage but does not necessarily implicate the marker site as directly contributing to the trait. The phenotypic differences between backcross genotypic marker classes depend on the effect of the quantitative trait locus (QTL) linked to the marker site and on the recombination fraction between the two²³. Although an interval

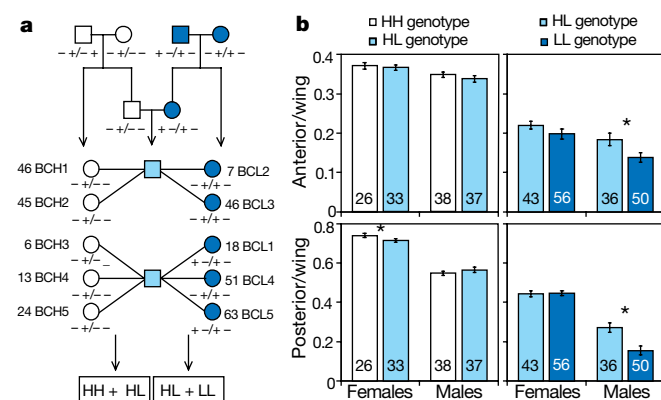


Figure 2 Test for an association between alleles at the *Dll* locus and dorsal eyespot size in *B. anynana*. **a**, Crosses between phenotypically divergent lines produced after nine generations of artificial selection. Individuals from the high and low lines and heterozygotes are represented by white, dark-blue and light-blue symbols, respectively. To minimize intra-line genetic variation we used single mating pairs (females shown as circles and males as squares). The backcrosses between two hybrid males (*Lepidoptera* females appear not to have recombination³⁰) and five females from each divergent line produced a total of five backcross families in each direction (BCH1–5 and BCL1–5 for high and low, respectively). The numbers before each BC family indicate total number of informative progeny. H and L represent high and low *Dll* alleles, respectively. *Dll* locus genotypes based on the two sequence polymorphisms are given for all single pairs used (see Methods; _ is not known). **b**, Mean eyespot/wing size index (± standard error) for males and females of each of the two genotypic classes (different colours) for each backcross direction (high line on the left and low line on the right). Asterisks indicate statistically significant differences at the 5% level (Table 1). Numbers inside columns represent sample sizes.

Table 1 Significant differences in eyespot size between *Dll* genotypes

Parameter	Female posterior (HH × HL)	Male anterior (HL × LL)	Male posterior (HL × LL)
Size difference*			
Trait values (T)	0.025	0.046	0.117
T/base s.d.	0.42	0.80	2.17
T/(high – low)	0.05	0.14	0.22
Probability of no eyespot†			
given HH	0	NA	NA
given HL	0	0.11	0.14
given LL	NA	0.24	0.48
P-value‡			
Mann–Whitney U	0.0132	0.0103	0.0004
Permutation test	0.001	0.007	0.004
Maximum LOD score			
Value	0.81	1.24	2.67
Position§ (cM)	20ll	0	0

NA, not applicable.

* Phenotypic differences between genotypic classes (HH, HL and LL) for the posterior eyespot of females from the high line backcross and for both eyespots of males from the low line. The differences between average phenotypes of alternative genotypes are given in raw trait units (T), units scaled to base population standard deviations (s.d.), and expressed as a proportion of the total difference between high and low selected lines.

† Probability of the eyespot being absent conditional on genotype.

‡ P-values for the one-way Mann–Whitney U-test and permutation tests indicate that all differences are statistically significant.

§ Distance from *Dll*.

|| LOD score at *Dll* is 0.73.

mapping approach is necessary to distinguish between the effect of a QTL and its location²⁴, we can calculate the log odds (LOD) score (see Methods) for a putative QTL some recombination fraction away from the markers in *Dll*. For both the posterior and anterior eyespots in males from the low backcross, the LOD score was maximized at *Dll* (Table 1), although a 1 LOD support interval indicates that the QTL could be as far as 40 cM from *Dll*. The effect detected for the posterior eyespot in females from the high backcross might be due to the same QTL that affects males (its LOD score did not peak at *Dll* but the value at *Dll* was not significantly lower than its maximum value; Table 1) or to a second QTL linked to, but not at *Dll*. The analyses in Table 1 indicate that a QTL at *Dll* is the most parsimonious explanation for the observed data, although a linked QTL cannot be ruled out. As *B. anynana* is thought to have approximately 26 linkage groups (26–29 is the haploid chromosome number in other *Bicyclus* species²⁵), linkage to *Dll* rules out a large number of other candidate genes. Mapping of the QTL to one of 26–29 linkage groups localizes it to about 4% of the genome. To improve this resolution and locate the causative sites within *Dll*, future work will use population-based association studies^{26,27}.

Dll codes for an evolutionarily conserved transcription factor characteristically expressed in insect appendages¹⁸. In butterflies it has been redeployed in a spatial coordinate system operating within specific wing regions and leading to eyespot formation^{6–8}. Our results suggest that this same gene contributes to inter-individual variation in eyespot size in *B. anynana*. The co-segregation of DNA polymorphisms in *Dll* with eyespot variation, coupled with quantitative changes in *Dll* expression between phenotypically divergent lines, suggests that not only the *Dll* pathway but molecular variants at *Dll* itself contribute to within-species variation in eyespot patterns and are thus likely to contribute to the evolutionary diversification of butterfly wing patterns. Other candidate genes possibly involved in generating variation in eyespot patterns include *decapentaplegic*⁶, *engrailed*^{9,11}, *spalt*¹¹ and *Ultrabithorax*¹⁰. Ultimately, we can ask how such genes interact with each other, the environment, and the pigment biochemical pathways to provide the evolutionary potential for the spectacular diversity seen in butterfly wing patterns. Our results show that a combination of approaches from evolutionary and developmental biology can contribute more to understanding how evolutionarily relevant variation is generated than either approach used alone.

Methods

Artificial selection

Nine generations of artificial selection on eyespot size targeted the ratios between eyespot diameter and a linear measurement of wing size. To select on the two dorsal forewing eyespots simultaneously (anterior, A, and posterior, P), the sum of the rank order of A/wing and P/wing was used ($rA + rP$). From an outbred laboratory stock³, high and low lines were derived by selecting for larger and smaller eyespots, respectively. To establish these lines, 1,058 females were measured and allowed to mate with males at random. The 45 females with most extreme $rA + rP$ values in each direction were used as parents for the next generation. In each of the following generations, 150–200 females per line were scored and the 30–40 most extreme selected.

Bicyclus anynana Dll

A *B. anynana* Dll homologue was obtained from an embryonic complementary DNA library using a probe from the *P. coenia* Dll gene (GenBank accession number AF404110). DNA sequence polymorphisms that distinguish high and low line Dll alleles were obtained by sequencing 1,250 bp of the 3' untranslated region (UTR) of *B. anynana* Dll (GenBank AF404825; bases 2,079–3,328) in three low and two high butterflies. For sequencing we used the same *B. anynana*-specific primers that were used to amplify the 3' UTR Dll fragment: 5'-ACTTATAAACACACAAATGGCGTA-3' and 5'-TTAAATATGTATCACAACAGACTGG-3'.

Genotyping assays

To type a C/T polymorphism (GenBank AF404825; base 2,194) located within the recognition sequence for the endonuclease *Bsp*HI, 20–100 ng genomic DNA were used in a 25- μ l polymerase chain reaction (PCR) with 2.5 U Taq polymerase in its recommended buffer (Pharmacia), 0.5 mM of each primer (5'-ACTTATAAACACACAAATGGCGTA-3' and 5'-ATGCTGTAATAATAATCGCATA-3') and 0.1 mM of each dNTP. Cycling conditions were 1 min at 96 °C followed by 35 cycles of 45 s at 94 °C, 45 s at 57 °C and 1 min at 72 °C. Amplicons were ethanol-precipitated, resuspended in a 20- μ l restriction digestion mix containing 5 U *Bsp*HI in its recommended buffer (New England Biolabs), incubated overnight at 37 °C, and scored on agarose gels. Individuals homozygous for the presence or absence of the cut site are scored +/+ and –/–, respectively; heterozygotes are scored +/-.

To type a 10-bp insertion-deletion polymorphism (GenBank AF404825; position 2,263), we designed two primers in the forward direction: 5'-GGTCACGGTATAAAA CGTGAACG-3' (insertion-specific) and 5'-GCCACGGTATAAACGTCAGT-3' (deletion-specific). These were used in separate PCRs with the same reverse primer (5'-GGGTGAACACAATTGCTGCT-3'). PCRs were carried out as above except for the annealing temperature (here 63 °C) and scored on agarose gels. Homozygotes for the insertion or the deletion (genotypes +/+ and –/–, respectively) have a band only for one of the primer combinations, whereas heterozygotes (+/-) have bands for both. Genotypes for both sites are given in the order of the polymorphisms along the cDNA sequence (*Bsp*HI, indel) with homologous chromosomes separated by a solidus (/).

These assays were used to genotype the butterflies from our crosses. The *Bsp*HI C/T polymorphism was fully informative in all backcross high (BCH) families. To genotype the low direction recombinants we used both assays. The *Bsp*HI polymorphism was fully informative in the backcross low families BCL1 and BCL5. For the BCL4 family, the two polymorphisms were used to provide an informative haplotype. For the families BCL2 and BCL3 only half of the progeny had informative genotypes for both polymorphisms; a total of 28 progeny of genotype –/+– were not included in the segregation analysis.

Genotype–phenotype association

Given that the phenotypes of the progeny from the BCL families did not follow a normal distribution, we have used a nonparametric statistical test (one-way Mann–Whitney *U*-test) to compare eyespot/wing size between different genotypic classes. The reason for applying one-tailed tests is the *a priori* expectation that the high and low alleles should, respectively, increase and decrease eyespot size. A permutation approach²⁸ was used to test the significance of the observed Mann–Whitney *U*-test results. Within each sex and backcross combination genotypes were randomly permuted 1,000 times with respect to paired eyespot size measures and the Mann–Whitney *U*-test statistic estimated. Distributions for the *n*th largest test statistic were then calculated and compared to the corresponding observed *n*th largest test statistic in the non-permuted data set. This approach allows *P*-values to be assigned to the most significant (second most significant, and so on) eyespot/sex/backcross combination, without correcting for potential correlations over characters, sexes, or backcross direction.

QTL position

To identify the most likely location of the QTL contributing to eyespot variation relative to the marker in *Dll* the following log likelihood equation was maximized for values of *r* between 0.0 and 0.5:

$$\log \text{likelihood} = \sum_i \begin{cases} \text{if } (y_i = 0) \{ \log[\theta_1(1-r) + \theta_2 r] + \log[\theta_1 r + \theta_2(1-r)] \} \\ \text{else } \{ \log[(1-\theta_1)(1-r)z(y_i, \mu, \sigma) + (1-\theta_2)r z(y_i, \mu + \beta, \sigma)] \\ \quad + \log[(1-\theta_1)r z(y_i, \mu, \sigma) + (1-\theta_2)(1-r)z(y_i, \mu + \beta, \sigma)] \} \end{cases}$$

where y_i is the phenotype of the *i*th butterfly, θ_1 is the probability (eyespot is absent; genotype is LL), θ_2 is the probability (eyespot is absent; genotype is HL), β is the expected

(eyespot size; eyespot is present and genotype is HL) minus the expected (eyespot size; eyespot is present and genotype is LL), $z(y_i, a, \sigma)$ is the normal probability distribution function with mean *a* and standard deviation σ evaluated at y_i , and *r* is the recombination fraction between the *Dll* marker and the putative QTL. Base ten LOD²³ scores were calculated for different values of *r* by maximizing the above function for $\theta_1, \theta_2, \mu, \beta, \sigma$ minus the same function maximized for the same parameters at *r* = 0.5 (unlinked QTL).

Dll expression patterns

Wing discs from late final instar larvae and 20–24 h pupa were stained with cross-reactive Dll antibody (1:100) as in ref. 8. As a secondary antibody we used donkey anti-rabbit Texas red (1:200; Jackson Laboratories). Images were collected on an MRC 1024 ES laser confocal microscope using a $\times 10$ objective for the entire larval wings and $\times 20$ for the eyespot centres.

Received 16 August; accepted 8 November 2001.

- Nijhout, H. F. *The Development and Evolution of Butterfly Wing Patterns* (Smithsonian Institution, Washington DC, 1991).
- Brakefield, P. M. The evolution–development interface and advances with the eyespot patterns of *Bicyclus* butterflies. *Heredity* **80**, 265–272 (1998).
- Monteiro, A. F., Brakefield, P. M. & French, V. The evolutionary genetics and developmental basis of wing pattern variation in the butterfly *Bicyclus anynana*. *Evolution* **48**, 1147–1157 (1994).
- Monteiro, A. F., Brakefield, P. M. & French, V. Butterfly eyespots: the genetics and development of the color rings. *Evolution* **51**, 1207–1216 (1997).
- Monteiro, A. F., Brakefield, P. M. & French, V. The genetics and development of an eyespot pattern in the butterfly *Bicyclus anynana*: response to selection for eyespot shape. *Genetics* **146**, 287–294 (1997).
- Carroll, S. B. *et al.* Pattern formation and eyespot determination in butterfly wings. *Science* **265**, 109–114 (1994).
- Carroll, S. B. Developmental regulatory mechanisms in the evolution of insect diversity. *Development* (Suppl.) 217–223 (1994).
- Brakefield, P. M. *et al.* Development, plasticity and evolution of butterfly eyespot patterns. *Nature* **384**, 236–242 (1996).
- Keys, D. N. *et al.* Recruitment of a *hedgehog* regulatory circuit in butterfly eyespot evolution. *Science* **283**, 532–534 (1999).
- Weatherbee, S. D. *et al.* *Ultrabithorax* function in butterfly wings and the evolution of insect wing patterns. *Curr. Biol.* **9**, 109–115 (1999).
- Brunetti, C. R. *et al.* The generation and diversification of butterfly eyespot color patterns. *Curr. Biol.* **11**, 1578–1585 (2001).
- Palopoli, M. F. & Patel, N. H. Neo-Darwinian developmental evolution: can we bridge the gap between pattern and process? *Curr. Opin. Gen. Dev.* **6**, 502–508 (1996).
- Stern, D. L. Perspective: evolutionary developmental biology and the problem of variation. *Evolution* **54**, 1079–1091 (2000).
- Mackay, T. F. C. Quantitative trait loci in *Drosophila*. *Nature Rev. Genet.* **2**, 11–20 (2001).
- Brakefield, P. M. & French, V. Butterfly wings: the evolution of development of colour patterns. *BioEssays* **21**, 391–401 (1999).
- Nijhout, H. F. Pattern formation on lepidopteran wings: determination of an eyespot. *Dev. Biol.* **80**, 267–274 (1980).
- French, V. & Brakefield, P. M. Eyespot development on butterfly wings: the focal signal. *Dev. Biol.* **168**, 112–123 (1995).
- Panganiban, G., Nagy, L. & Carroll, S. B. The role of the *Distal-less* gene in the development and evolution of insect limbs. *Curr. Biol.* **4**, 671–675 (1994).
- Kreitman, M. Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. *Nature* **304**, 412–417 (1983).
- Long, A. D. *et al.* High resolution mapping of genetic factors affecting abdominal bristle number in *Drosophila melanogaster*. *Genetics* **139**, 1273–1291 (1995).
- Taylor, B. A., Wnek, C., Schroeder, D. & Phillips, S. J. Multiple obesity QTLs identified in an intercross between the NZO (New Zealand obese) and the SM (small) mouse strains. *Mamm. Genome* **12**, 95–103 (2001).
- Wijngaarden, P. J. & Brakefield, P. M. The genetic basis of eyespot size in the butterfly *Bicyclus anynana*: an analysis of line crosses. *Heredity* **85**, 471–479 (2000).
- Edwards, M. D., Stuber, C. W. & Wendel, J. F. Molecular-marker-facilitated investigations of quantitative-trait loci in maize. I. Numbers, genomic distribution and types of gene action. *Genetics* **116**, 113–125 (1987).
- Lander, E. S. & Botstein, D. Mapping Mendelian factors underlying quantitative traits using RFLP Linkage maps. *Genetics* **121**, 185–199 (1989).
- Robinson, R. *Lepidoptera Genetics* (Pergamon, Oxford, 1971).
- Lai, C. G., Lyman, R. F., Long, A. D., Langley, C. H. & Mackay, T. F. C. Naturally-occurring variation in bristle number and DNA polymorphisms at the *scabrous* locus of *Drosophila melanogaster*. *Science* **266**, 1697–1702 (1994).
- Long, A. D., Lyman, R. F., Morgan, A. H., Langley, C. H. & Mackay, T. F. C. Both naturally occurring insertions of transposable elements and intermediary frequency polymorphisms at the *achaete-scute* complex are associated with variation in bristle number in *Drosophila melanogaster*. *Genetics* **154**, 1255–1269 (2000).
- Doerge, R. W. & Churchill, G. A. Permutation test for multiple loci affecting a quantitative character. *Genetics* **142**, 285–294 (1996).
- Falconer, D. S. & Mackay, T. F. C. *Introduction to Quantitative Genetics* 4th edn (Addison Wesley Longman, Harlow, Essex, 1996).
- Marec, F. Synaptonemal complexes in insects. *Int. J. Insect Morphol. Embryol.* **25**, 205–233 (1996).

Acknowledgements

We thank K. Kooops for rearing the butterflies, and B. Zwaan for help with the crosses; C. Brunetti, J. Selegue and S. Carroll for the *B. anynana* cDNA library, the *P. coenia* Dll probe and the Dll antibody; N. Glansdorp for help with wing extractions, and G. Lammers

for help with the confocal microscopy; K. Broman for deriving the likelihood function; F. Marec for information on butterfly chromosomes; M. Brittijn for help with the figures; and A. Monteiro, H. Teotônio and B. Zwaan for comments on the manuscript. This work was supported by the Portuguese Foundation for Science and Technology and the Luso-American Foundation under the Gulbenkian Program (P.B.), the Human Frontier Science Program Organization (P.M.B.) and the US National Institute of Health (A.D.L.).

Correspondence and requests for materials should be addressed to P.B. (e-mail: pbeldade@rulsfb.leidenuniv.nl).

Visual categorization shapes feature selectivity in the primate temporal cortex

Natasha Sigala & Nikos K. Logothetis

Max Planck Institute for Biological Cybernetics, Spemannstrasse 38, 72076 Tuebingen, Germany

The way that we perceive and interact with objects depends on our previous experience with them. For example, a bird expert is more likely to recognize a bird as a sparrow, a sandpiper or a cockatiel than a non-expert¹. Neurons in the inferior temporal cortex have been shown to be important in the representation of visual objects; however, it is unknown which object features are represented and how these representations are affected by categorization training. Here we show that feature selectivity in the macaque inferior temporal cortex is shaped by categorization of objects on the basis of their visual features. Specifically, we recorded from

single neurons while monkeys performed a categorization task with two sets of parametric stimuli. Each stimulus set consisted of four varying features, but only two of the four were important for the categorization task (diagnostic features). We found enhanced neuronal representation of the diagnostic features relative to the non-diagnostic ones. These findings demonstrate that stimulus features important for categorization are instantiated in the activity of single units (neurons) in the primate inferior temporal cortex.

Perception is arguably the end product of a categorization process². A number of studies have tried to elucidate the ways that features are extracted and represented by non-human primates in the context of a categorization task^{3–10}. These studies have shown that rhesus monkeys are capable of representing natural objects at basic (for example trees versus non-trees)^{5,6,8}, superordinate (animals or food)^{2,3,7} and subordinate (individual wire objects, faces or fish)^{9,10} levels. However, it still remains to be shown which object features are represented and how these representations are affected by the categorization training. The inferior temporal cortex area has a critical role in visual object recognition and responds to complex stimuli^{9,11} as well as to greatly simplified versions of these stimuli¹². Moreover, activity in the human temporal cortex is thought to be sensitive to the categorization level of the stimuli and to depend on the expertise of the observer^{13,14}. The aim of our study was to test whether inferior temporal cortex neurons respond selectively to object features that constitute the relevant dimensions for visual object categorization.

We trained two monkeys in a categorization task and used parameterized line drawings of faces and fish as stimuli (Fig. 1). Each schematic face consisted of an outline and four varying features: eye height (EH), eye separation (ES), nose length (NL) and mouth height (MH)¹⁵. The schematic fish also had four features: the shape of the dorsal fin (DF), tail (T), ventral fins (VF) and mouth (M). The four-dimensional vectors of parameters

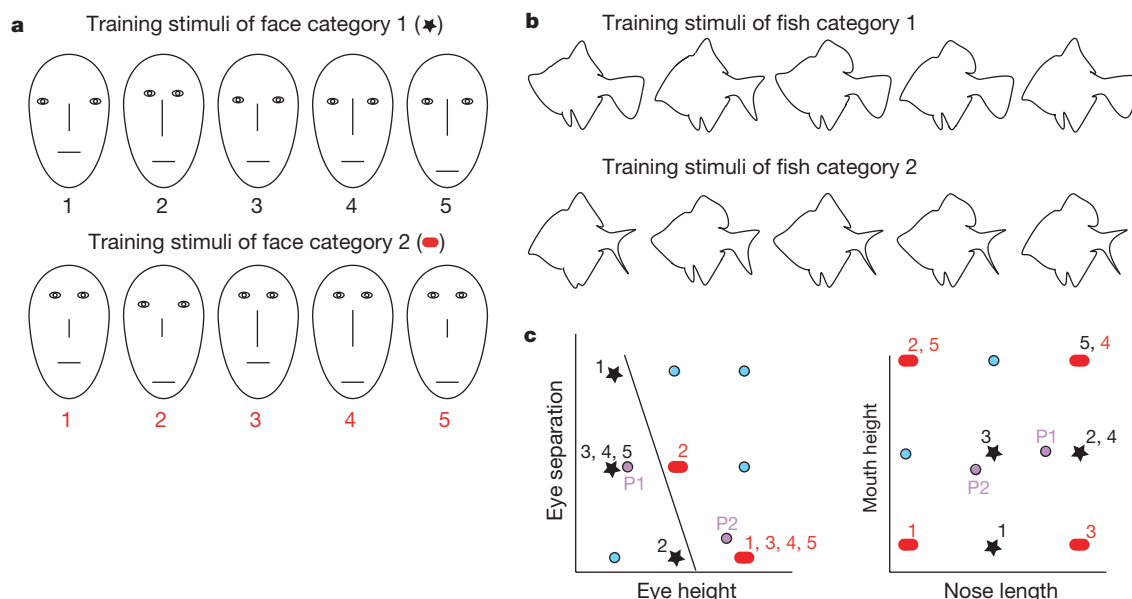


Figure 1 Stimuli and categories. **a**, The first stimulus set consisted of line drawings of faces with four varying features: eye height, eye separation, nose length and mouth height. **b**, The second stimulus set consisted of fish outlines with four varying features: the shape of the dorsal fin, tail, ventral fins and mouth. In both stimulus sets, each feature could take one of three discrete values. The categories were separable along two of the four stimulus features, but information about only one of the diagnostic features was insufficient for optimal performance. The monkeys were presented with one stimulus at a time. **c**, Two-dimensional representation of the stimulus space. Black stars represent the

stimuli of the first category and red ovals represent the stimuli of the second category. Each number indicates the position of one corresponding stimulus from **a**. As the stimuli differ along four dimensions, the two-dimensional representations in this figure result in overlap of stimuli that are distinct. The purple circles (P1 and P2) represent the prototypes. Cyan circles represent test exemplars that did not belong to a fixed category. The two categories were linearly separable along the eye height, eye separation (EH, ES) dimensions (solid line) but not along the nose length, mouth height (NL, MH) dimensions.

characterizing each stimulus are shown in Fig. 2 by their projections onto two two-dimensional plots. Each feature could take three discrete values and the categories were separable along two of the four dimensions of the stimuli. Specifically, the faces were separable along the (EH, ES), but not along the (NL, MH) dimensions (Fig. 2, solid line); the fish outlines were separable along (DF, T), but not along (VF, M) dimensions. The red circles in Fig. 1 correspond to the training exemplars of one category and the black stars to the training exemplars of the other category. The two purple circles correspond to the prototype of each category, formed by averaging the feature values of the five corresponding training exemplars. The cyan circles illustrate positions of test exemplars chosen to span a large area of the stimulus space between and around the training exemplars. The monkeys were presented with one stimulus at a time and they were required to pull one lever after they saw a stimulus from category one and to pull a different lever for the stimuli from category two. At the time of the combined psychophysical–electrophysiological experiments the performance and median reaction times were 98% correct and 416 ms, respectively, for monkey A and 98% correct and 530 ms, respectively, for monkey B.

We isolated 150 single units (neurons) from the anterior inferior temporal cortex, 96 of which showed significant visual responses to the line drawings of faces and fish and to other visual stimuli that included line drawings and natural images (one-tailed *t*-test versus baseline activity, $P < 0.05$). All 96 visually responsive units were tested with the face stimuli ($n = 34$), and 65 of them were also tested with the fish stimuli ($n = 34$), in all cases with 10 repetitions per stimulus. To test the effect of categorization training on the shaping of feature selectivity, we grouped the responses of the neurons to all the stimuli of one stimulus set and performed a one-way analysis of variance (ANOVA) for each one of the four stimulus features. Forty-four out of 96 neurons (45.8%) responded significantly differently to at least one feature value of the face stimuli (one-way ANOVA, $P < 0.05$ corrected for the four comparisons). Twenty-eight out of 65 neurons (43.1%) responded significantly differently to at least

one feature value of the fish stimuli. Of this population of feature-selective neurons, 72.7% (32/44) were selective for one or both diagnostic features of the face stimuli (and not for the non-diagnostic features), and 75% (21/28) were selective for one or both diagnostic features of the fish stimuli (and not for the non-diagnostic features). Figure 2 shows an example of a neuron whose firing rate discriminates different values of one diagnostic feature of the EH stimuli, while it does not distinguish the different values of the non-diagnostic features. Figure 3 shows the population average for all the neurons tested with the face stimuli ($n = 96$).

To compare directly selectivity for diagnostic versus non-diagnostic features, we examined the entire population of visually responsive neurons, irrespective of whether they showed feature selectivity. We computed a standard selectivity index for each of the four features (the difference between maximum and minimum activity for a given feature divided by their sum). Figure 4 shows the average selectivity index of each neuron for the diagnostic compared with the non-diagnostic features. If the neurons were equally selective for diagnostic and non-diagnostic features, the data points should cluster around the diagonal (dotted line). Instead, the average selectivity index for the diagnostic features is significantly higher than for the non-diagnostic ones ($P < 10^{-8}$ for the faces; $P < 10^{-4}$ for the fish). This indicates more precise tuning for the features that were diagnostic for the categorization task compared with the features that were not diagnostic. These results show that although inferior temporal cortex neurons were responsive to all the parameterized stimuli irrespective of category, and to many different control stimuli, they nonetheless carried detailed, parametric information for the most behaviourally relevant stimulus features.

Recently we reported that monkeys and humans follow the same categorization strategies for the stimulus sets used in the present study¹⁰. Moreover, we showed that categorization training

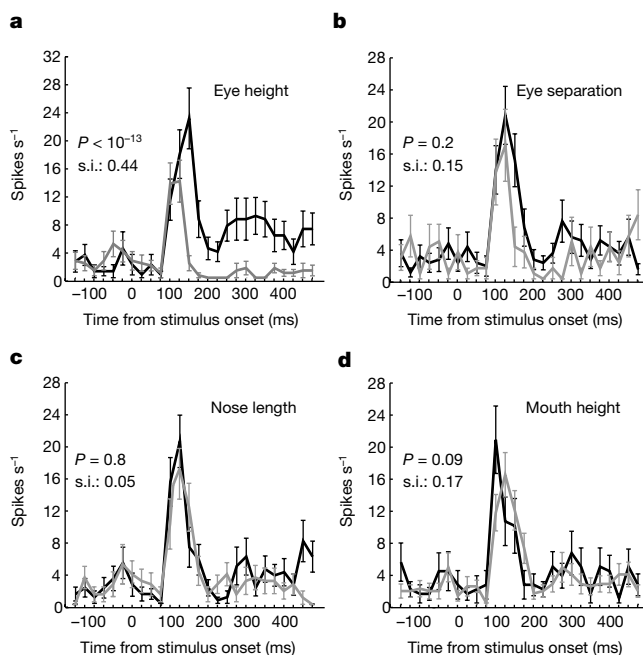


Figure 2 Example of a neuron showing feature selectivity. The average activity is sorted by stimulus feature (eye height (a), eye separation (b), nose length (c) and mouth height (d)). Black traces indicate the responses of the neuron to the best feature value; grey traces indicate responses to the worst feature value. The bars indicate standard error of the mean. s.i., selectivity index (see text).

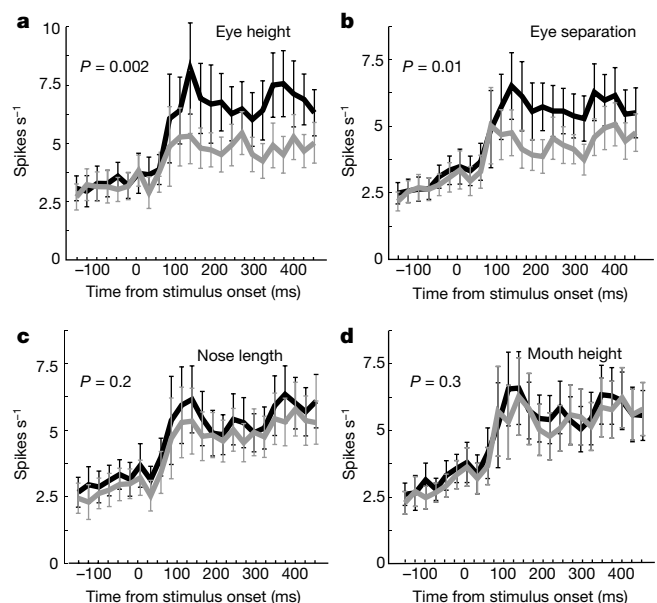


Figure 3 Population average for the neurons tested with the face stimuli. The mean firing rate of every neuron was normalized to its background activity in the 200 ms period before stimulus onset. Black traces indicate average responses to the best feature value; grey traces indicate average responses to the worst feature value. The bars indicate standard error of the mean. For each feature (a–d) a *t*-test was performed for the time window 100–475 ms after stimulus presentation, testing the hypothesis that the mean firing rate to the best feature was higher than the mean firing rate to the worst feature. The hypothesis could not be rejected in the case of the diagnostic features (eye height (a) and eye separation (b)).

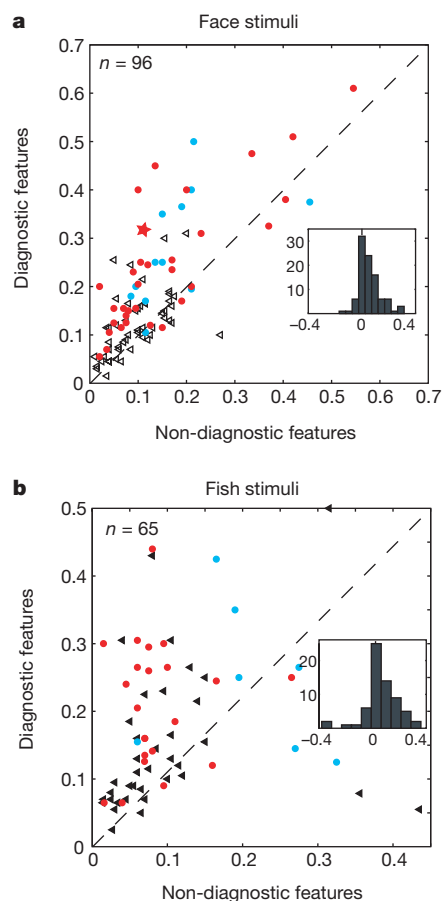


Figure 4 Representation of the diagnostic features in the neuronal population for faces (**a**) and fish (**b**). Plots of the average selectivity index of each neuron for the diagnostic versus the non-diagnostic features. Each point represents one neuron. Red circles represent neurons with statistically significant selectivity for diagnostic dimensions only. Blue circles represent neurons with statistically significant selectivity for diagnostic and non-diagnostic features. Black triangles represent neurons with no significant selectivity. The red star represents the example neuron depicted in Fig. 2. The dotted line indicates equal selectivity for all stimulus features. Inserts show the distribution (numbers of neurons) of the selectivity index for the diagnostic features after subtraction of the selectivity index for the non-diagnostic features.

affects the way that the stimuli are perceived in the context of a different task. The data from the present study indicate that neuronal selectivity was shaped by the most relevant subset of features during the categorization training. Furthermore, we suggest that this neuronal fine-tuning reflects the monkeys' perceptual increase of sensitivity for the diagnostic features, as revealed in the psychophysical part of our study¹⁰. Our finding that inferior temporal cortex neurons can selectively represent visual object features that are important for the task at hand provides a mechanism through which feature diagnosticity shapes the encoding and the perceptual interpretation of visual objects¹⁹. □

Methods

Behavioural and electrophysiological methods

Two adult male monkeys (*Macaca mulatta*) participated in the experiments. All studies were approved by the local authorities and were in full compliance with the guidelines of the European Community (EUV, European Union directive 86/609/EEC) for the care and use of laboratory animals. The stimuli were presented on a 21-inch monitor placed at a distance of 97 cm from the monkeys. The angular size subtended by each face line drawing was $2.4^\circ \times 4.4^\circ$ and the angular size of the fish line drawings was $3.5^\circ \times 2.4^\circ$. The monkeys maintained gaze within 4° of the centre of the stimulus. Eye movements were recorded by virtue of the scleral search coil technique²⁰ and sampled at 200 Hz (CNC Engineering).

The monkeys were trained to perform the categorization task following standard operant conditioning techniques with positive reinforcement. They were seated in a custom-made primate chair with two response levers mounted on the front. During the learning phase the monkeys were presented with the ten training exemplars in random order. Five of them belonged to category 1 and were assigned to one lever, and five of them belonged to category 2 and were assigned to the other lever. After the monkeys reached a minimum of 75% correct, they were trained with blocks of presentations and received a juice reward after five consecutive correct responses. The feedback for incorrect responses was an auditory signal and a delay of 2 s before the next stimulus presentation. After the monkeys reached a performance level of 85% correct for the training exemplars, they were presented with these ten exemplars as controls, interleaved with 24 new stimuli. The monkeys received feedback for incorrect categorization of the training exemplars only. This procedure was followed separately for both stimulus sets (faces and fish).

Single-cell activity was recorded for both monkeys in the inferior temporal cortex using a recording chamber consisting of a ball-and-socket joint with an 18-gauge stainless steel tube passing through its centre^{16,17}. We recorded from two hemispheres (one left and one right) in two monkeys with the chambers placed at AP = 20, L = 21 (AP, anterior–posterior; L, lateral). The recording sites, as estimated from the stereotaxic coordinates of the guide tube, magnetic resonance imaging (MRI) images and white to grey matter transitions during the recordings were in temporal areas TEm, TE1 and TE2. Our recording methodology has been described in more detail elsewhere¹⁸. During recordings the monkeys were categorizing 34 stimuli of one stimulus set.

Data analysis

We analysed activity in the 300 ms period after the onset of the visual response. We defined the onset of the visual response as the time when the activity was equal to or higher than two standard deviations of the background activity in the 200 ms period before stimulus onset. In the trials that we included in our analysis the monkey had not pressed a lever up to that point.

Received 12 October; accepted 9 November 2001.

1. Tanaka, J. W. & Taylor, M. Object categories and expertise: is the basic level in the eye of the beholder? *Cogn. Psychol.* **23**, 457–482 (1991).
2. Bruner, J. S. On perceptual readiness. *Psychol. Rev.* **64**, 123–152 (1957).
3. Sands, S. E., Lincoln, C. E. & Wright, A. A. Pictorial similarity judgments and the organization of visual memory in the rhesus monkey. *J. Exp. Psychol. Gen.* **111**, 369–389 (1982).
4. Sugihara, T., Edelman, S. & Tanaka, K. Representation of objective similarity among three-dimensional shapes in the monkey. *Biol. Cybern.* **78**, 1–7 (1998).
5. Vogels, R. Categorization of complex visual images by rhesus monkeys. Part 1: behavioural study. *Eur. J. Neurosci.* **11**, 1223–1238 (1999).
6. Vogels, R. Categorization of complex visual images by rhesus monkeys. Part 2: single-cell study. *Eur. J. Neurosci.* **11**, 1239–1255 (1999).
7. Delorme, A., Richard, G. & Fabre-Thorpe, M. Ultra-rapid categorisation of natural scenes does not rely on colour cues: a study in monkeys and humans. *Vision Res.* **40**, 2187–2200 (2000).
8. Freedman, D. J., Riesenhuber, M., Poggio, T. & Miller, E. K. Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* **291**, 312–316 (2001).
9. Logothetis, N. K., Pauls, J. & Poggio, T. Shape representation in the inferior temporal cortex of monkeys. *Curr. Biol.* **5**, 552–563 (1995).
10. Sigala, N., Gabbiani, F. & Logothetis, N. K. Visual categorization and object representation in monkeys and humans. *J. Cogn. Neurosci.* **14**, 1–12 (2001).
11. Gross, C. G., Roche-Miranda, C. E. & Bender, D. B. Visual properties of neurons in the inferotemporal cortex of the monkey. *J. Neurophysiol.* **35**, 96–111 (1972).
12. Tanaka, K. Inferotemporal cortex and object vision. *Annu. Rev. Neurosci.* **19**, 109–139 (1996).
13. Gauthier, I., Anderson, A. W., Tarr, M. J., Skudlarski, P. & Gore, J. C. Levels of categorization in visual recognition studied using functional magnetic resonance imaging. *Curr. Biol.* **7**, 645–651 (1997).
14. Gauthier, I., Skudlarski, P., Gore, J. C. & Anderson, A. W. Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neurosci.* **3**, 191–197 (2000).
15. Nosofsky, R. M. Attention, similarity, and the identification–categorization relationship. *J. Exp. Psychol. Gen.* **115**, 39–57 (1986).
16. Schiller, P. H. & Koerner, F. Discharge characteristics of single units in superior colliculus of the alert rhesus monkey. *J. Neurophysiol.* **34**, 920–936 (1971).
17. Sheinberg, D. L. & Logothetis, N. K. The role of temporal cortical areas in perceptual organization. *Proc. Natl Acad. Sci. USA* **94**, 3408–3413 (1997).
18. Sheinberg, D. L. & Logothetis, N. K. Noticing familiar objects in real world scenes: the role of temporal cortical neurons in natural vision. *J. Neurosci.* **21**, 1340–1350 (2001).
19. Schyns, P. Diagnostic recognition: task constraints, object information, and their interactions. *Cognition* **67**, 147–179 (1998).
20. Judge, S., Richmond, B. & Chu, F. Implantation of magnetic search coils for measurement of eye position: An improved method. *Vision Res.* **20**, 535–538 (1980).

Acknowledgements

We thank Z. Kourtzi, M. Silver, S. Smirnakis and A. Tolias for comments on the manuscript and discussions. We also thank J. Pauls and D. Sheinberg for help and advice at various stages of the project. This work was supported by the Max Planck Society.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to N.S. (e-mail: natasha.sigala@tuebingen.mpg.de).

RIM1 α forms a protein scaffold for regulating neurotransmitter release at the active zone

Susanne Schoch^{*†}, Pablo E. Castillo^{†‡§}, Tobias Joll, Konark Mukherjee^{*}, Martin Geppert[¶], Yun Wang^{*}, Frank Schmitz^{*¶}, Robert C. Malenka[‡] & Thomas C. Südhof^{*}

^{*} The Center for Basic Neuroscience and Department of Molecular Genetics, Howard Hughes Medical Institute, The University of Texas Southwestern Medical Center, Dallas, Texas 75390-9111, USA

[‡] Nancy Friend Pritzker Laboratory, Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, California 94304, USA

[§] Department of Neuroscience, Albert Einstein College of Medicine, Bronx, New York, New York 10461, USA

[¶] Department of Molecular Neurobiology, Max-Planck-Institut für experimentelle Medizin, 37075 Göttingen, Germany

[†] These authors contributed equally to this work

[¶] Present addresses: Genomics Sciences, Pharmacia Corporation, Kalamazoo, Michigan 49007, USA (M.G.); Leibniz Institut für Neurobiologie, 39118 Magdeburg, Germany (F.S.)

Neurotransmitters are released by synaptic vesicle fusion at the active zone^{1,2}. The active zone of a synapse mediates Ca²⁺-triggered neurotransmitter release, and integrates presynaptic signals in regulating this release. Much is known about the structure of active zones and synaptic vesicles, but the functional relation between their components is poorly understood³. Here we show that RIM1 α , an active zone protein that was identified as a putative effector for the synaptic vesicle protein Rab3A^{4,5}, interacts with several active zone molecules, including Munc13-1 (ref. 6) and α -liprins^{7,8}, to form a protein scaffold in the presynaptic nerve terminal. Abolishing the expression of RIM1 α in mice shows that RIM1 α is essential for maintaining normal probability of neurotransmitter release, and for regulating release during short-term synaptic plasticity. These data indicate that RIM1 α has a central function in integrating active zone proteins and synaptic vesicles into a molecular scaffold that controls neurotransmitter release.

Morphological data suggest that a protein scaffold organizes the active zone and links it physically to synaptic vesicles^{3,9}. Four active zone proteins have been identified that may function as synaptic scaffolds: RIMs (RIM1 α and RIM2 α) are putative effectors for Rab3A, a synaptic vesicle protein that regulates synaptic vesicle exocytosis^{4,5,10}; Unc13/Munc13s are required for priming synaptic vesicles for exocytosis^{11–13}; α -liprins bind to receptor tyrosine phosphatases⁷ and are essential for normal active zone function in *Caenorhabditis elegans*⁸; and RIM-BPs form tight complexes with RIMs⁵. Vertebrate synapses also contain large presynaptic proteins called bassoon and piccolo/aczonin, which are not evolutionarily conserved and are absent from some synapses, such as neuromuscular junctions³.

How these active zone proteins contribute to the coordination of presynaptic signals is unknown. RIMs engage in several interactions that might guide active zone function. For example, RIMs bind to Rab3A, which might link presynaptic active zones containing RIMs to docked synaptic vesicles containing Rab3A^{4,5}, although the functional importance of this interaction has been questioned^{14,15}. RIMs also bind to RIM-BPs through a carboxy-terminal sequence⁵, to cyclic-AMP-GEFII through an unknown site¹⁶, to Ca²⁺ channels and synaptotagmin 1 through both C₂ domains¹⁷, and to Munc13-1 through their amino-terminal zinc-finger^{18,19}. The *in vivo* relevance of these interactions and the function of RIMs at the active zone are unknown.

To test the synaptic function of RIMs in mice, we abolished expression of RIM1 α , the full-length transcript of the RIM1 gene and most abundant RIM isoform^{4,5}. We isolated murine genomic clones containing the first coding exon of the RIM1 α gene, and constructed a targeting vector that replaces this exon with a neomycin resistance gene (Supplementary Information). Because the first exon includes the initiator methionine and 55 conserved N-terminal residues of RIM1 α ⁴, deleting this exon should abolish expression. After homologous recombination in embryonic stem cells, clones carrying the targeted RIM1 α allele were used to produce mutant mice. All analyses were performed on littermates derived from mating heterozygous mutant mice on a hybrid SV129/black 6 background, and were confirmed with several independent litters derived from many generations of heterozygous breedings.

Homozygous RIM1 α mutant mice were viable and fertile. The number of offspring from heterozygous matings surviving into adulthood (2 months and older) showed a slightly skewed ratio disfavouring homozygous mutants (25% wild-type, 55% heterozygous, and 20% homozygous; $n = 695$). Mutant mice were noticeable as they did not care for their litters even after several pregnancies, suggesting a deficit in maternal behaviour (data not shown). However, morphological analyses of the mutant brains failed to uncover structural abnormalities or changes in brain architecture (Supplementary Information).

Immunoblotting using antibodies to the N-terminal zinc-fingers of RIM1 α and RIM2 α confirmed that the RIM1 α mutation abolished expression of RIM1 α , but had no effect on RIM2 α (Fig. 1a). The RIM2 α antibodies cross-reacted with RIM1 α , but the RIM1 α antibody did not cross-react with RIM2 α . Because RIM1 α and RIM2 α have slightly different sizes on immunoblots, we could estimate the relative abundance of RIM1 α and RIM2 α in forebrain with the RIM2 α antibodies. RIM1 α was at least twice as abundant as RIM2 α . Thus, by targeting RIM1 α , we deleted the main RIM isoform in the knockout mice. Both RIM antibodies also cross-reacted with rabphilin (Fig. 1a), an unexpected observation given the distant similarity between the RIM and rabphilin zinc-finger domains⁴. However, rabphilin was unchanged in RIM1 α knockout mice, as were RIMs in rabphilin knockout mice (Fig. 1a).

We surveyed several synaptic proteins to determine whether deleting RIM1 α had an effect on the overall composition of the brain. Of 30 proteins tested, only Munc13-1 showed a change in expression (Fig. 1b, c), decreasing by about 60% in the knockout mouse presumably because Munc13-1 binds to RIM1 α ^{18,19} and

Table 1 Quantitative ultrastructural analysis of CA1 synapses

Parameter	Wild-type mice	RIM1 α knockout mice
Synapse density (per 10 ³ μ m ²)	315 $\pm$ 112	302 $\pm$ 86
Presynaptic terminal area (per 10 ⁻² μ m ²)	22.4 $\pm$ 11.7	23.5 $\pm$ 11.9
Synaptic vesicle cluster density (per 10 ³ μ m ²)	6.94 $\pm$ 1.45	6.78 $\pm$ 1.38
Synaptic vesicle density in terminals (N per μ m ² nerve terminal area)	245 $\pm$ 157	205 $\pm$ 126
Size of the active zone (nm)	234 $\pm$ 90	261 $\pm$ 90
Docked synaptic vesicles (per μ m ² active zone)	16.2 $\pm$ 5.2	16.7 $\pm$ 6.5

All data are from apical dendrites in the distal stratum radiatum of the hippocampal CA1 region. Analyses were performed on randomized masked electron micrographs to avoid observer bias. All data shown are the mean $\pm$ s.d. determined from 160 synapses of 4 independent animals of each genotype, except for the synapse density, which was calculated from 62 synapses from 4 animals.

becomes destabilized in its absence, similar to the destabilization of rabphilin in the absence of Rab3A²⁰. There are two additional isoforms of Munc13-1 in brain: Munc13-2, which is expressed widely; and Munc13-3, which is observed primarily in the cerebellum⁶. The decrease of Munc13-1 in the RIM1 α knockouts was selective for Munc13-1; Munc13-2 (which does not bind to RIM1 α) showed no decrease, and Munc13-3 was only moderately affected (Supplementary Information).

No significant changes were observed in Rab3A, synaptotagmin 1 or RIM-BP expression, suggesting that not all proteins that bind to RIM1 α are altered in the knockout mice. The concentration of several postsynaptic density proteins was increased moderately (Fig. 1c; and Supplementary Information). The coordinate increase of GRIP, SynGAP, PSD95 and SHANK suggested that deleting RIM1 α might lead to synaptic remodelling; however, electron micrographs from various regions of the hippocampus revealed no obvious abnormalities (data not shown). We searched for quantitative changes with an unbiased morphometric analysis of hippocampal CA1 region synapses in the distal stratum radiatum of wild-type and mutant mice. We detected no statistically significant difference in the density and size of synapses, synaptic vesicle density, or number of docked vesicles (Table 1), indicating that the RIM1 α deletion does not lead to an overall remodelling of synapses. We observed a small, but not significant, increase in active zone length, as found to a greater extent in the *C. elegans* α -liprin mutation⁷.

To analyse the function of RIM1 α in synaptic transmission, we recorded synaptic responses at excitatory synapses formed by Schaffer collateral/commissural fibres on CA1 pyramidal cells. Because RIM1 α was identified as a Rab3A effector⁴, we directly compared RIM1 α and Rab3A knockout mice¹⁰. In addition, to test whether the changes observed in RIM1 α knockout mice depend on the reduced expression of Munc13-1 that occurs in these mice (Fig. 1c), we examined heterozygous Munc13-1 knockout mice that also contain decreased concentrations of Munc13-1. All experiments were performed blind, with direct comparison of littermates derived from heterozygous matings.

Presynaptic stimulation triggered a robust postsynaptic response in RIM1 α and Rab3A knockouts and in Munc13-1 heterozygous

mice, consistent with the fact that these mutations are not lethal. To search for regulatory abnormalities, we examined paired-pulse facilitation (PPF), which is the enhancement of transmitter release in response to two closely spaced stimuli. PPF is normally inversely correlated with the probability of neurotransmitter release (P_r), such that synapses with low P_r show large PPF, whereas synapses with high P_r show small PPF or even paired-pulse depression (PPD)²¹. Rab3A knockout mice show a selective increase in PPF at short interstimulus intervals (ISIs) without a change in P_r (ref. 10), suggesting that Rab3A may function in regulating neurotransmitter release. RIM1 α - and the Rab3A-deficient hippocampal slices showed a similarly large increase in PPF at short ISIs (<50 ms) but not at longer ISIs (>100 ms; Fig. 2a, b). In contrast, Munc13-1 heterozygous mutant mice produced no change in PPF (Fig. 2c). Thus, in this important form of short-term synaptic plasticity, RIM1 α - and Rab3A-deficient synapses displayed similar abnormalities that were not caused by a reduction in concentrations of Munc13-1.

Little is known about the proteins that control GABA (γ -aminobutyric acid) release at inhibitory synapses. Analysis of synapses lacking Munc13-1 has indicated that the mechanisms controlling vesicle exocytosis at excitatory and inhibitory synapses may be markedly different¹³. To determine whether GABA release is regulated normally in RIM1 α knockout mice, we recorded monosynaptic inhibitory postsynaptic currents (IPSCs) from CA1 pyramidal neurons in response to paired-pulse stimuli in the stratum radiatum. Inhibitory synapses generally express PPD in response to paired stimuli, probably because they have a high P_r (ref. 22). Unexpectedly, PPD was increased at short ISIs in RIM1 α knockout mice (Fig. 2d). Therefore, deleting RIM1 α decreases transmitter release during paired stimulations at inhibitory synapses, opposite to the increase at excitatory synapses (Fig. 2a).

This abnormality at inhibitory synapses in RIM1 α knockouts might be a result of the high P_r of inhibitory synapses. We therefore re-analysed inhibitory synapses after decreasing the P_r by lowering the Ca^{2+} concentration (1 mM Ca^{2+} , 5 mM Mg^{2+}), which converts PPD at inhibitory synapses into PPF. Consistent with the results obtained at normal Ca^{2+} concentrations, however, PPF at inhibitory

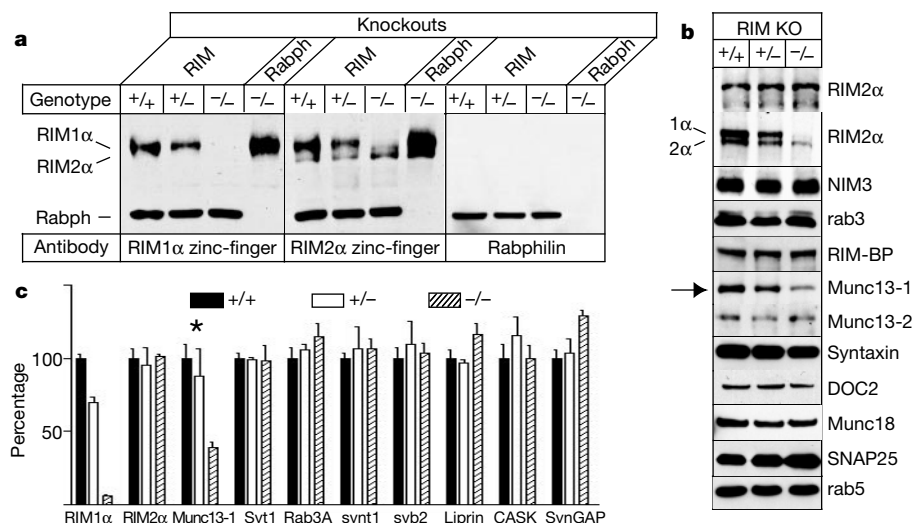


Figure 1 Generation and protein analysis of RIM1 α knockout mice. **a**, Immunoblotting analysis of RIM and rabphilin expression in the brains of wild-type (+/+), heterozygous (+/-) and homozygous (-/-) RIM1 α and rabphilin (Rabph) knockout mice. Blots were analysed with antibodies to the N-terminal zinc-finger domains of RIM1 α (Q703) and RIM2 α (U1565), and to rabphilin (734). **b**, Representative immunoblots of the levels of synaptic proteins in brains from wild-type, heterozygous and homozygous RIM1 α knockout mice. One of the two RIM2 α antibodies used (indicated on left) cross-reacts with

RIM1 α ; arrow marks Munc13-1. **c**, Relative levels of synaptic proteins in wild-type, heterozygous, and homozygous mutant RIM1 α knockout mice as quantified by immunoblotting with ¹²⁵I-labelled secondary antibodies and phosphorimager detection. All signals are normalized for internal standards probed on the same blots²⁸ (see Supplementary Information for a larger survey of proteins). Note the large change in Munc13-1 levels (asterisk).

synapses under decreased P_r was reduced in RIM1 α knockout mice (Fig. 2d). These results indicate that, similar to Munc13-1 (ref. 13), RIM1 α may have distinct functions in excitatory and inhibitory synapses. We also examined PPD at inhibitory synapses in Rab3A knockout mice. Unlike PPF at excitatory synapses in these mice, PPD at physiological concentrations of $\text{Ca}^{2+}/\text{Mg}^{2+}$ was nearly normal, as was PPF under conditions of decreased P_r (Fig. 2e). These results indicate that RIM1 α acts mostly independently of Rab3A in inhibitory synapses.

We also examined longer-lasting forms of short-term synaptic plasticity. During repetitive stimulation of excitatory synapses at moderate frequencies (such as 14 Hz), control synaptic responses exhibited initial facilitation that slowly decreased. By contrast, RIM1 α - and Rab3A-deficient synapses, but not synapses from Munc13-1 heterozygotes, displayed similar extended facilitation, consistent with increased PPF at short ISIs (Fig. 3a–c). Next, we tested post-tetanic potentiation (PTP), a longer lasting form of short-term plasticity. PTP elicited by 30 stimuli at 100 Hz (in the presence of AP5 (D-2-amino-5-phosphonopentanoate) was strongly enhanced in the RIM1 α knockout mice (Fig. 3d), but

unaffected in the Rab3A knockout mice and the Munc13-1 heterozygotes (Fig. 3e, f). This indicates that RIM1 α has a role in the regulation of glutamate release independent of its interaction with Rab3A as described above for inhibitory synapses. We also examined LTP, which was induced normally in the CA1 region of the RIM1 α knockout mice (Supplementary Information), suggesting that in contrast to mossy fibre LTP²³, NMDA (N-methyl-D-aspartate) receptor (NMDAR)-dependent LTP does not require RIM1 α .

The changes in PPF, PTP and facilitation at excitatory synapses in RIM1 α knockout mice are consistent with an overall decrease in P_r . As a rough estimate of the overall synaptic strength in wild-type and mutant mice, we constructed synaptic ‘input–output’ curves. A decrease in the magnitude of the synaptic responses in RIM1 α knockout mice was observed (Fig. 4a). Because this result is consistent with a decrease in P_r but does not prove it, we measured P_r directly using the MK-801 method^{24,25}. The inhibitor MK-801 irreversibly blocks NMDARs, but only when they are activated by synaptically released glutamate.

Repetitive stimulation in the presence of MK-801 led to a

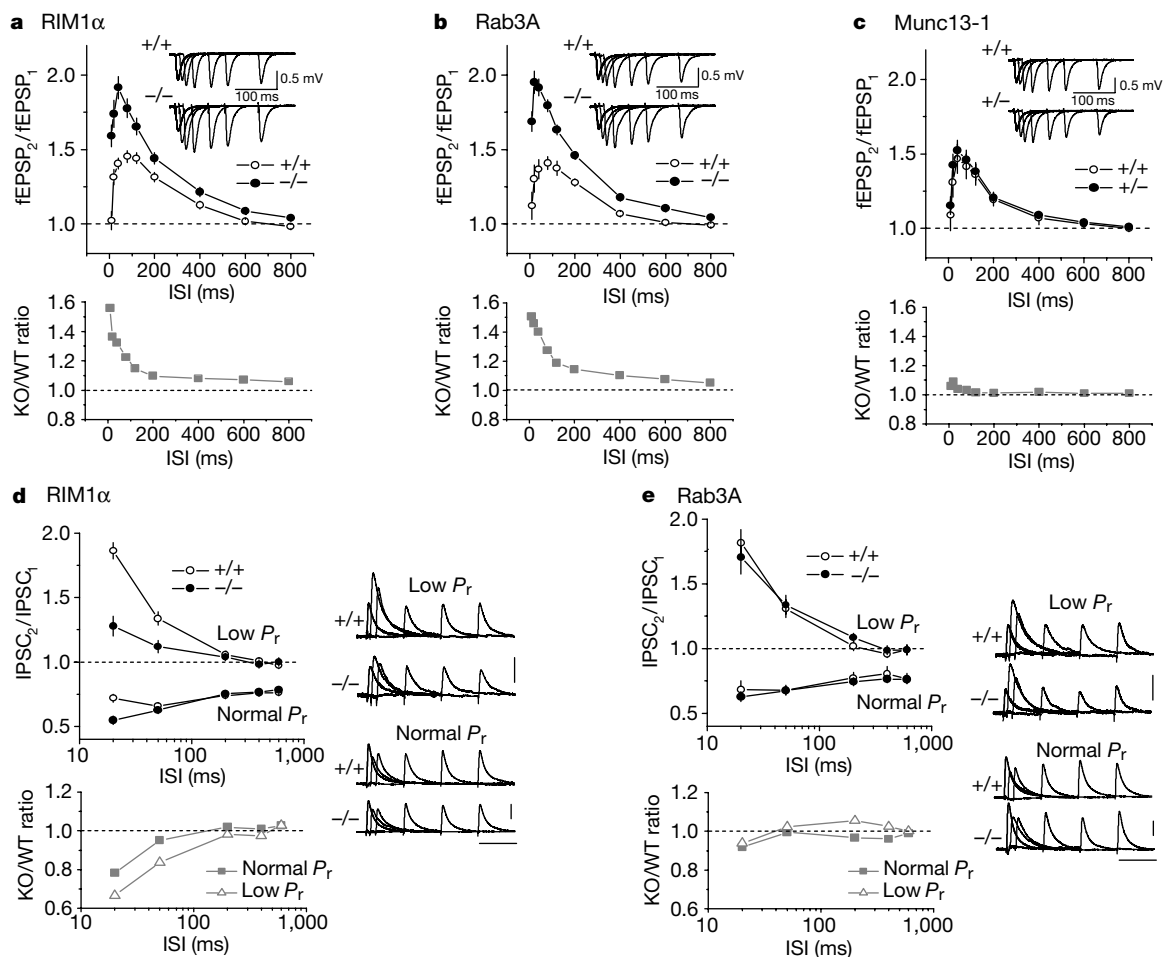


Figure 2 Excitatory and inhibitory synaptic responses to paired-pulse stimulation in mutant mice. **a–c**, Paired-pulse facilitation ($fEPSP_2/fEPSP_1$) recorded as a function of the interstimulus interval (ISI) at excitatory Schaffer collateral/CA1 pyramidal cell synapses (top graphs in each panel). Bottom graphs in each panel display the ratio of the paired-pulse facilitation values between wild-type and mutant mice. Insets show typical fEPSP averages of 10–20 consecutive responses elicited at different ISIs from wild-type (+/+) and mutant (–/–) mice. Responses are superimposed after subtraction of the first response. Data are from wild-type and RIM1 α knockout (**a**) or Rab3A knockout mice (**b**) and from wild-type and Munc13-1 heterozygous mice (**c**). Data are from the following

number of mice/slices: knockout 4/14, wild type 4/15 (**a**); knockout 3/7, wild type 2/6 (**b**); knockout 3/8, wild type 3/9 (**c**). **d, e**, Paired-pulse responses at inhibitory synapses in the CA1 region of the hippocampus in RIM1 α (**d**) or Rab3A knockout mice (**e**). The ratio of inhibitory synaptic currents ($IPSC_2/IPSC_1$) was measured as a function of the ISI in extracellular solution containing 3 mM Ca^{2+} , 3 mM Mg^{2+} (normal P_r), or 1 mM Ca^{2+} , 5 mM Mg^{2+} (low P_r). Graphs and traces as in **a–c**. IPSC calibration bars, 200 ms and 100 pA (low P_r) or 200 pA (normal P_r). Data are from the following number of mice/slices: knockout 4/9, wild type 4/8 (**d**); knockout 2/6, wild type 2/6 (**e**).

continuous decrease in NMDAR-mediated excitatory postsynaptic currents (EPSCs), with the rate of decrease being directly proportional to P_r (refs. 24, 25). After application of MK-801, NMDAR-mediated EPSCs declined significantly slower in RIM1 α knockout mice than in wild-type controls (Fig. 4b). When fitted to a single exponential as a rough measure of the average P_r , the difference in the decay constants indicated that the P_r in the mutant slices is roughly half of that in control slices, in agreement with the input–output curves (Fig. 4a, b).

In parallel, we also examined P_r in Rab3A knockout mice using the MK-801 method (Fig. 4c). Consistent with previous results¹⁰, we detected no difference between Rab3A-deficient and wild-type mice, providing further evidence that RIM1 α regulates glutamate release at excitatory synapses independently of Rab3A. There is no analogous technique to estimate P_r at inhibitory synapses. We therefore used the relatively indirect assay of comparing the frequency of miniature inhibitory postsynaptic currents (mIPSCs) in wild-type and RIM1 α knockout mice, but found no difference indicative of a change in P_r between the two populations (wild-type, 19.5 ± 8.2 Hz, $n = 6$ cells/3 mice; knockout, 18.8 ± 9.3 Hz, $n = 7$ cells/2 mice).

The electrophysiological phenotype suggests that RIM1 α per-

forms a multifaceted role in synapses that can be divided into Rab3A-dependent components (the changes in PPF, synaptic run-down and mossy fibre LTP; Figs 2 and 3, and ref. 23) and Rab3A-independent components (the changes in PTP, synaptic strength and P_r ; Figs 3 and 4). A possible explanation for the Rab3A-independent phenotype of the RIM1 α knockout is provided by the decrease in Munc13-1 (which acts as a priming factor in exocytosis¹³); however, synapses lacking Munc13-1, or with less Munc13-1, did not show any of the characteristic features of the RIM1 α knockout (Figs 2 and 3, and ref. 13). Thus, the function of RIM1 α as an integrator of synaptic signals, as inferred from the phenotype of the knockout, may involve other interactions with active zone proteins. We therefore searched systematically for proteins that interact with RIMs using yeast two-hybrid screens and GST pull-downs. In addition to Rab3A we identified three interacting proteins (Fig. 5): Munc13-1 and synaptotagmin 1, which have been shown to bind to RIM1 α ^{17–19}; and α -liprins, adaptor proteins that are essential for the presynaptic active zone in *C. elegans*⁸.

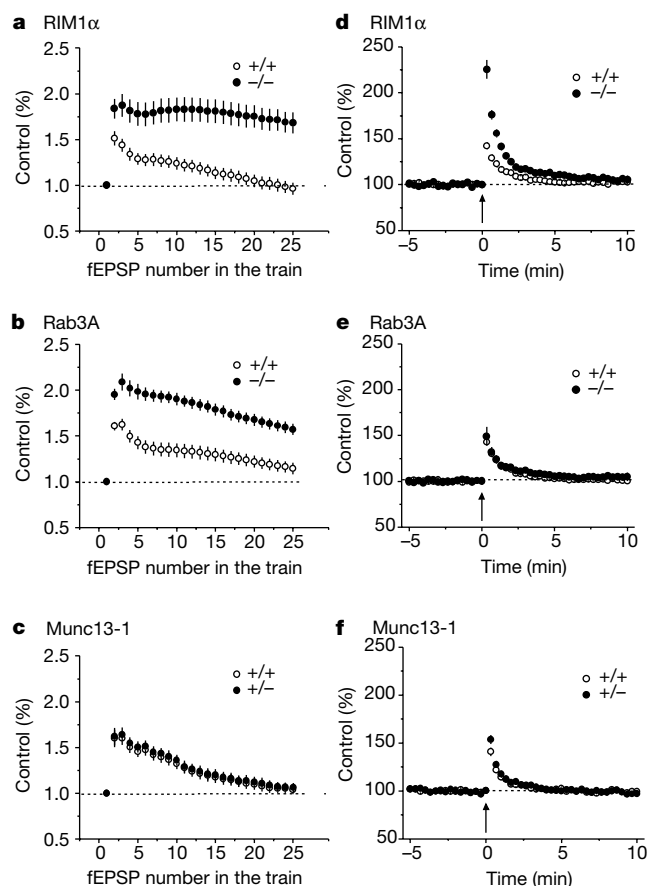


Figure 3 Synaptic plasticity in RIM1 α and Rab3A knockout mice. **a–c**, Response to 14-Hz stimulus trains at Schaffer collateral/CA1 pyramidal cell synapses in RIM1 α knockout (5 mice/11 slices) and wild-type (5/10) mice (**a**), in Rab3A knockout (3/7) and wild-type (2/6) mice (**b**), and in Munc13-1 heterozygous (3/8) and wild-type (3/9) mice (**c**). Values are normalized to the first synaptic response of the stimulation train. **d–f**, Time course of post-tetanic potentiation induced by a 100-Hz, 30-pulse tetanus (arrow) in presence of $40 \mu\text{M}$ AP5 in RIM1 α knockout (3 mice/12 slices) and wild-type (3/10) mice (**d**); Rab3A knockout (2/5) and wild-type (2/6) mice (**e**); and Munc13-1 heterozygous (3/8) and wild-type (3/9) mice (**f**).

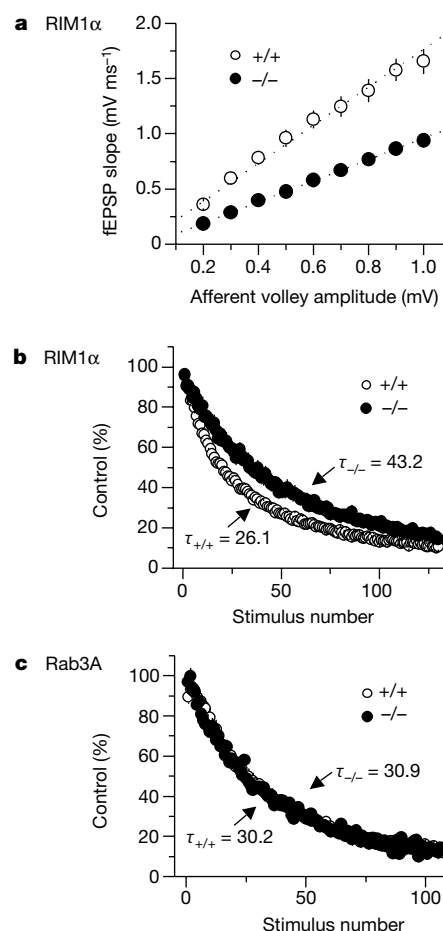


Figure 4 Synaptic efficacy and release probability at Schaffer collateral/CA1 pyramidal cell synapses in RIM1 α and Rab3A knockout mice. **a**, Analysis of synaptic strength using input–output measurements. The extracellular field potential slope (fEPSP slope) is plotted as a function of the afferent volley amplitude. Linear fit slopes for wild-type (1.72 ± 0.07 ; $r = 0.99$; 3 mice/10 slices) and RIM1 α knockout mice (0.96 ± 0.01 ; $r = 0.99$; 3 mice/9 slices) are significantly different ($P < 0.002$, unpaired *t*-test). **b, c**, Time course of MK-801-dependent blockage of NMDAR-mediated EPSC amplitude (normalized to baseline) at 0.1-Hz stimulation. The rate of blockade was fitted by single exponentials (RIM1 α knockout, $\tau_{-/-} = 43.2 \pm 1.0$ (16 cells/5 mice); wild type, $\tau_{+/+} = 26.1 \pm 0.5$ (18 cells/5 mice); Rab3A knockout, $\tau_{-/-} = 29.3 \pm 1.3$ (7 cells/3 mice); wild type, $\tau_{+/+} = 30.5 \pm 1.2$ (6 cells/2 mice)).

The identification of α -liprins as RIM interactors in yeast two-hybrid screens was the most unexpected finding (Fig. 5a). More than 50% of all clones isolated encoded liprin- α 3 or liprin- α 4. All clones contained the N-terminal α -liprin sequence corresponding to residues 200–350 of liprin- α 1 (ref. 7), with variable amounts of flanking sequences. An interaction between RIMs with liprins is potentially interesting, because RIM is mislocalized in *C. elegans* liprin mutants (Y. Jin and M. Nonet, personal communication).

In pull-down experiments, GST fusion proteins containing the C₂B domain of RIMs bound to α -liprins (Fig. 5b, d), and reverse experiments with α -liprin–GST fusion proteins confirmed this

binding (data not shown). As many C₂ domains bind Ca²⁺, these experiments were performed in both the presence and absence of Ca²⁺ ions. Ca²⁺ had no effect on liprin binding, as expected from the lack of Ca²⁺-binding consensus sequences in RIM C₂ domains. We also confirmed reports^{18,19} that Munc13-1 binds to the N-terminal zinc-finger of RIM1 α (data not shown). In contrast to previous findings¹⁷, however, we found that synaptotagmin 1 bound only to the C₂B domain and not to the C₂A domain of RIMs, and binding was tightly dependent on Ca²⁺ (Fig. 5c, d). Because the C₂A and C₂B domains of RIMs are only distantly related, such differential binding is not unexpected. These data show that RIMs interact through distinct sequences with several synaptic proteins that have diverse but essential functions in neurotransmitter release (Fig. 5e).

We have shown that RIM1 α participates in the control of transmitter release at both excitatory and inhibitory synapses at a late stage in the vesicle cycle, presumably during vesicle priming and fusion. On the basis of the profound physiological changes at RIM1 α -deficient synapses and the biochemical interactions of RIMs with presynaptic proteins, we propose that RIM1 α , and by extension also RIM2 α , function at the active zone to integrate the actions of several molecules with diverse roles in neurotransmitter release. Specifically, we suggest that RIMs form a macromolecular complex through interactions with three active zone proteins (Munc13, RIM-BP and α -liprin) and two synaptic vesicle proteins (Rab3A and synaptotagmin 1; Fig. 5e). As a result, RIMs link proteins with known functions in synaptic vesicle priming (Munc13-1) and organizing active zones (α -liprins) to synaptic vesicle proteins essential for regulating synaptic vesicle fusion (Rab3A and synaptotagmin 1). The finding that RIM1 α binds to all known evolutionarily conserved components of active zones is striking. Although the physiological changes observed in RIM1 α knockout mice are severe, they cannot be attributed to a major structural alteration of the active zone, or to a significant redistribution of synaptic vesicles. This suggests either that the molecular scaffold formed by interactions of RIM1 α with other active zone proteins functions primarily in signalling, or that the RIM2 α remaining in the knockout mice, although of lower abundance, is sufficient to maintain the overall active zone structure.

How might the protein–protein interactions of RIM1 α account for the two main changes in presynaptic function observed in the RIM1 α knockout mice: the abnormalities in short-term synaptic plasticity and the decrease in P_r at excitatory synapses? Synapses lacking Rab3A show a selective increase in PPF in the absence of a change in P_r . Because the RIM1 α knockout mimicked the Rab3A knockout with regard to PPF (but not to PTP and P_r), it seems plausible that the enhancement of PPF is caused by the loss of the Rab3A–RIM1 α interaction. Given the speed with which the change in PPF sets in (<10 ms), mechanical coupling of the Rab3A–RIM1 α complex might be a possible explanation for this change. Notably, however, the RIM1 α knockout also showed changes at excitatory and inhibitory synapses not present in the Rab3A knockout, indicating that the functions of RIM1 α go beyond those attributable to Rab3A, and that the protein machinery controlling GABA release at inhibitory synapses is markedly different from that used for glutamate release. This conclusion is supported by the opposite effects of the RIM1 α deletion on excitatory and inhibitory synapses. Although the interactions of RIM1 α with α -liprins or RIM-BPs might be responsible for the change in P_r , the most attractive candidate for mediating this effect is synaptotagmin 1 because this protein interacts with RIMs in a Ca²⁺-dependent manner and is a key regulator of P_r (ref. 26). Similar to the effects of mutating synaptotagmin 1, the change in P_r in the RIM1 α knockouts does not seem to involve a change in the number or proportion of docked vesicles, and thus probably reflects a modification of the release apparatus.

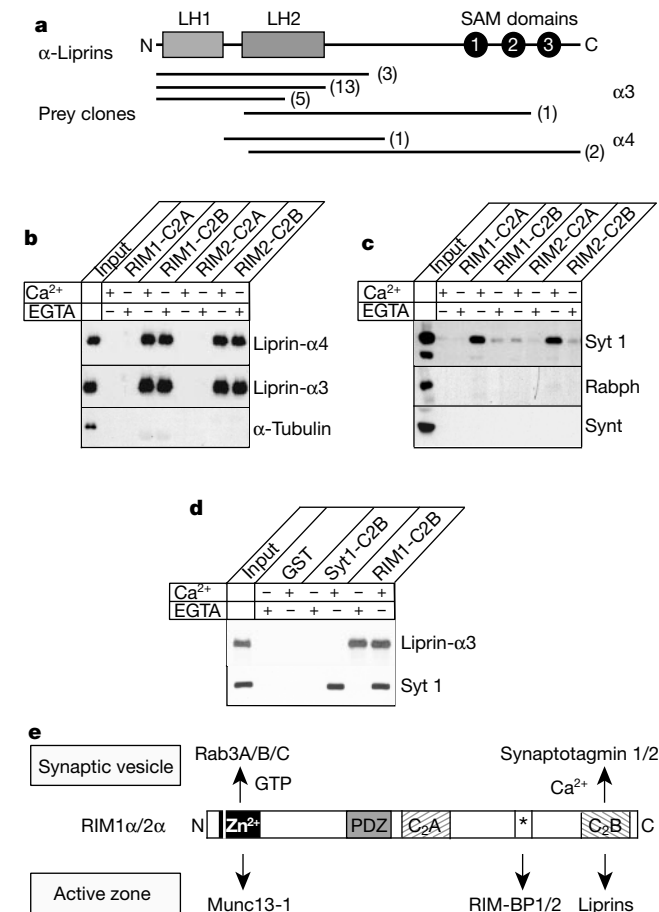


Figure 5 Biochemical definition of RIM1 α as a synaptic scaffolding protein. **a**, α -Liprins isolated by yeast two-hybrid screens with RIM1 α C₂-domains. The domain structure of α -liprins with two α -liprin homology regions at the N terminus (LH1 and LH2) and three C-terminal SAM domains (circles) is shown above the locations of the six independent liprin- α 3 and - α 4 prey clones, which accounted for more than 50% of all preys isolated. Number of isolates for each prey clone is shown in parentheses. **b**, GST pull-downs of recombinant α -liprins (Myc-tagged N-terminal regions of liprin- α 3 and - α 4 produced in transfected COS cells) with the C₂A or C₂B domains of RIM1 α or RIM2 α in the presence and absence of Ca²⁺. Samples were analysed by immunoblotting, with α -tubulin as a negative control. **c**, GST pull-downs of rat brain synaptotagmin 1 with the RIM1 α and RIM2 α C₂A and C₂B domains in the presence and absence of Ca²⁺. Captured proteins were visualized with antibodies to synaptotagmin 1 (Syt 1), rabphilin (Rabph) and syntaxin 1 (Synt). **d**, Comparison of the Ca²⁺-dependent binding of brain liprin- α 3 and synaptotagmin 1 to the C₂B-domains from synaptotagmin 1 and RIM1 α analysed by GST pull-downs. **e**, Overview of protein–protein interactions mediated by RIMs. The domain structure of RIMs incorporates an N-terminal zinc-finger, central PDZ and C-terminal C₂ domains. Up arrows indicate GTP- or Ca²⁺-dependent interactions of RIMs with synaptic vesicle proteins; down arrows constitutive interactions with active zone proteins.

Methods

GST–protein affinity chromatography

Experiments were performed as described^{4,5,20} with proteins from brain or transfected COS cells solubilized by Triton X-100 or cholic acid. Bound proteins were analysed by SDS–polyacrylamide gel electrophoresis (PAGE) and Coomassie blue staining, followed by immunoblotting or protein sequencing.

Yeast two-hybrid methods

A postnatal day 8 (P8) rat brain complementary DNA library in pVP16-3 was screened by standard yeast two-hybrid methods^{4,5} with pLexN-RIM1C2 bait vector (residues 713–1588). Of 48 sequenced positive clones, 25 encoded overlapping sequences from the N terminus of liprin- α 3 or α 4 (GenBank accession codes AY057064 and AY057065). We confirmed all positive clones with several bait vectors from both RIMs; interactions were quantified for selected bait–prey pairs using liquid β -galactosidase assays^{4,5}.

RIM1 α knockout mice

A targeting vector constructed from a genomic λ clone (16-1) in pTK-Neo3A (ref. 27) was used for homologous recombination in E14.1 cells. The resulting chimaeric mice were monitored by coat colour and polymerase chain reaction (PCR) with the following primers: wild-type 1774, 5'-GGGAGGCCAGAGCATCATGAGAGACTG-3', and 1773, 5'-GCTTCACTTGGCGCTGCATATCTCAC-3', product size 0.7 kilobases (kb); knockout 1808, 5'-CAACTGTGGCTGTGCACACTTGGCCG-3', and 676, 5'-GAGCGCGC-GCGGCGGAGTTGTTGAC-3', product size 0.3 kb). Mouse husbandry was done as described^{27,28}; all analyses were carried out with littermates of heterozygote matings.

Protein analyses

Aliquots (40 μ g protein per lane) of brain proteins from three littermate mice (8 weeks old) per genotype were analysed by SDS–PAGE and quantitative immunoblotting^{27,28} with ¹²⁵I-labelled secondary antibodies and phosphorimager detection, using internal standards probed on the same blots for signal normalization^{27,28}.

Morphological studies

Anaesthetized adult mice (~2 months old) were perfusion-fixed with 4% fresh paraformaldehyde (light microscopy), or 2.5% glutaraldehyde, 2% paraformaldehyde, 0.1 M sodium cacodylate buffer (pH 7.2) pre-equilibrated to 37°C (electron microscopy), and examined by standard procedures²⁸. We quantified synaptic parameters on anonymized digital electron micrographs by the following criteria: active zone length is the presynaptic plasma membrane directly opposing the postsynaptic density; number of docked vesicles is all vesicles within 25 nm of the active zone; vesicle cluster density is the number of vesicles in a randomly pasted 80 $\times$ 80 nm square on the presynaptic vesicle cluster next to the active zone; density of synaptic vesicles is the total number of vesicles divided by the presynaptic terminal area²⁸.

Electrophysiological analysis

Synaptic responses were recorded from hippocampal slices (0.4 mm) of 3–6-week-old mice at room temperature by extracellular field potentials (fEPSPs) in the stratum radiatum in CA1, or by whole-cell voltage clamp from CA1 pyramidal cells^{23,27,28}. Unless otherwise stated, the external solution contained (in mM): 119 NaCl, 2.5 KCl, 1.2 MgSO₄, 2.5 CaCl₂, 1 NaHPO₄, 26.2 NaHCO₃, 10 glucose saturated with 95% O₂ and 5% CO₂ (pH 7.4). For extracellular recordings, the patch pipette contained 1 M NaCl, and initial slopes of fEPSPs were measured. Schaffer collaterals were stimulated at 0.1 Hz with a patch pipette filled with external solution. We adjusted the stimulation strength so that the baseline synaptic amplitude was constant among all experiments (fEPSPs, 0.5–1.0 mV; IPSCs, 200–300 pA; NMDAR-dependent EPSCs, 200–300 pA). Data were digitized (2 kHz) and analysed on-line using custom software for IgorPro (WaveMetrics). Synaptic depression was examined using 25 pulses at 14 Hz with 60-s intervals; 5 stimuli were averaged for each experiment. PTP was elicited by 30 stimuli at 100 Hz in 40 μ M AP5; LTP by four stimulus trains of 1 s at 100 Hz with 20-s intervals. NMDAR-mediated EPSCs were monitored in 10 μ M NBQX, 100 μ M picrotoxin at a holding potential of +30 mV. For assessing the rate of blockade of NMDAR-dependent EPSCs, a stable baseline of synaptic responses was established during 15 min stimulation at 0.1 Hz. MK-801 (25 μ M) was then bath applied without stimulation for 8 min, 0.1-Hz stimulation was then resumed, and the NMDAR-dependent blockade of EPSCs was determined. IPSCs were elicited by stimulation of the stratum radiatum and recorded in whole-cell configuration from CA1 pyramidal cells (V_h = +10 mV) in 10 μ M NBQX, 5 μ M CGP55845 and 50 μ M AP5, with the pipette solution containing (in mM): 123 caesium gluconate, 10 CsCl, 8 NaCl, 1 CaCl₂, 10 EGTA, 10 HEPES, 10 glucose (pH 7.2; 280–290 mOsm). Series resistance (ranging from 10 to 20 M Ω) was monitored throughout each experiment with a –4 mV, 80-ms pulse before every afferent stimulus; cells with more than 15% change in series resistance were excluded from analysis. All data were acquired on littermate offspring from heterozygous matings, and analysed without knowledge of the genotype of the tissue being studied. We subsequently confirmed the genotypes of individual mice by PCR. Chemicals were from Sigma except for MK-801, CGP55845 and AP5 (all from Tocris Cookson).

Miscellaneous

Most antibodies have been described^{15,12,27,28} or were obtained from Affinity BioReagents or Transduction Labs. Rabbit RIM1 α antibodies were raised against residues 11–399 (Q703),

492–722 (R809) and 1443–1455 (T2796); RIM2 α antibodies against residues 461–987 (U952), 159–319 (U957), 1336–1614 (U1130) and 1–466 (U1565); and liprin- α 3 antibodies (4396) against residues 184–470. We performed RNA blotting experiments using total RNA prepared from littermates of various genotypes, with 15 μ g RNA per lane and high-stringency hybridization with a probe from the RIM1 zinc-finger domain (residues 11–399), the C₂A-domain (residues 748–907), or the whole coding region of NIM3.

Received 26 September; accepted 5 November 2001.

- Katz, B. *The Release of Neural Transmitter Substances* (Liverpool Univ. Press, Liverpool, 1969).
- Cowan, W. M., Südhof, T. C. & Stevens, C. F. (eds) *Synapses* (Johns Hopkins Univ. Press, Baltimore, 2001).
- Garner, C. C., Kindler, S. & Gundelfinger, E. D. Molecular determinants of presynaptic active zones. *Curr. Opin. Neurobiol.* **10**, 321–327 (2000).
- Wang, Y., Okamoto, M., Schmitz, F., Hofman, K. & Südhof, T. C. RIM is a putative Rab3A-effector in regulating synaptic vesicle fusion. *Nature* **388**, 593–598 (1997).
- Wang, Y., Sugita, S. & Südhof, T. C. The RIM/NIM family of neuronal C₂-domain proteins: interactions with Rab3 and a new class of SH3-domain proteins. *J. Biol. Chem.* **275**, 20033–20044 (2000).
- Brose, N., Hofmann, K., Hata, Y. & Südhof, T. C. Mammalian homologues of *C. elegans* unc-13 gene define novel family of C₂-domain proteins. *J. Biol. Chem.* **270**, 25273–25280 (1995).
- Serra-Pages, C., Medley, Q. G., Tang, M., Hart, A. & Streuli, M. Liprins, a family of LAR transmembrane protein-tyrosine phosphatase-interacting proteins. *J. Biol. Chem.* **273**, 15611–15620 (1998).
- Zhen, M. & Jin, Y. The liprin protein SYD-2 regulates the differentiation of presynaptic termini in *C. elegans*. *Nature* **401**, 371–375 (1999).
- Harlow, M. L., Rens, D., Stoschek, A., Marshall, R. M. & McMahan, U. J. The architecture of active zone material at the frog's neuromuscular junction. *Nature* **409**, 479–484 (2001).
- Geppert, M., Goda, Y., Stevens, C. F. & Südhof, T. C. The small GTP-binding protein Rab3A regulates a late step in synaptic vesicle fusion. *Nature* **387**, 810–814 (1997).
- Richmond, J. E., Davis, W. S. & Jorgensen, E. M. UNC-13 is required for synaptic vesicle fusion in *C. elegans*. *Nature Neurosci.* **2**, 959–964 (1999).
- Aravamudan, B., Fergestad, T., Davis, W. S., Rodesch, C. K. & Broadie, K. *Drosophila* UNC-13 is essential for synaptic transmission. *Nature Neurosci.* **2**, 965–971 (1999).
- Augustin, I., Rosenmund, C., Südhof, T. C. & Brose, N. Munc13-1 is essential for fusion competence of glutamatergic synaptic vesicles. *Nature* **400**, 457–461 (1999).
- Coppola, T. et al. Disruption of Rab3-caldmodulin interaction, but not other effector interactions, prevents Rab3 inhibition of exocytosis. *EMBO J.* **18**, 5885–5891 (1999).
- Sun, L., Bittner, M. A. & Holz, R. W. Rab3A-binding and secretion-enhancing domains in Rim1 are separate and unique: Studies in adrenal chromaffin cells. *J. Biol. Chem.* **276**, 12911–12917 (2001).
- Ozaki, N. et al. cAMP-GEFII is a direct target of cAMP in regulated exocytosis. *Nature Cell Biol.* **2**, 805–811 (2000).
- Coppola, T. et al. Direct interaction of the rab3 effector rim with Ca²⁺ channels, snap-25, and synaptotagmin. *J. Biol. Chem.* **276**, 32756–32762 (2001).
- Betz, A. et al. Functional interaction of the active zone proteins Munc13-1 and RIM1 in synaptic vesicle priming. *Neuron* **30**, 183–196 (2001).
- Wang, X., Hu, B., Zimmermann, B. & Kilimann, M. W. Rim1 and rabphilin-3 bind Rab3-GTP by composite determinants partially related through N-terminal α -helix motifs. *J. Biol. Chem.* **276**, 32480–32488 (2001).
- Li, C. et al. Synaptic targeting of rabphilin-3A, a synaptic vesicle Ca²⁺/phospholipid-binding protein, depends on Rab3A/3C. *Neuron* **13**, 885–898 (1994).
- Thomson, A. M. Facilitation, augmentation and potentiation at central synapses. *Trends Neurosci.* **23**, 305–312 (2000).
- Kraushaar, U. & Jonas, P. Efficacy and stability of quantal GABA release at a hippocampal interneuron-principal neuron synapse. *J. Neurosci.* **20**, 5594–5607 (2000).
- Castillo, P. E., Schoch, S., Schmitz, F., Südhof, T. C. & Malenka, R. C. RIM1 α is required for presynaptic long-term potentiation. *Nature* **415**, 327–330 (2002).
- Rosenmund, C., Clements, J. D. & Westbrook, G. L. Nonuniform probability of glutamate release at a hippocampal synapse. *Science* **262**, 754–757 (1993).
- Hessler, N. A., Shirke, A. M. & Malinow, R. The probability of transmitter release at a mammalian central synapse. *Nature* **366**, 569–572 (1993).
- Fernandez-Chacon, R. et al. Synaptotagmin I functions as a calcium regulator of release probability. *Nature* **410**, 41–49 (2001).
- Schlüter, O. M. et al. Rabphilin knock-out mice reveal that rabphilin is not required for rab3 function in regulating neurotransmitter release. *J. Neurosci.* **19**, 5834–5846 (1999).
- Rosahl, T. W. et al. Essential functions of synapsins I and II in synaptic vesicle regulation. *Nature* **375**, 488–493 (1995).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com>).

Acknowledgements

We thank H. Riedesel for help with the RIM1 α knockout mice; R. Jahn and N. Brose for antibodies; and N. Brose, Y. Jin and M. Nonet for discussing results before publication. We are grateful to I. Leznicki, A. Roth and E. Borowicz for technical assistance. This study was supported by a fellowship to S.S. from the Deutscher Akademische Austauschdienst and grants from the NIH (R.C.M.).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.C.S. (e-mail: Thomas.Sudhof@UTSouthwestern.edu).

RIM1 α is required for presynaptic long-term potentiation

Pablo E. Castillo*^{†‡}, Susanne Schoch[§], Frank Schmitz^{||}, Thomas C. Südhof[§] & Robert C. Malenka*

* Nancy Friend Pritzker Laboratory, Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, California 94304, USA

[†] Department of Neuroscience, Albert Einstein College of Medicine, Bronx, New York 10461, USA

[§] The Center for Basic Neuroscience and Department of Molecular Genetics, Howard Hughes Medical Institute, The University of Texas Southwestern Medical Center, Dallas, Texas 75390-9111, USA

[‡] These authors contributed equally to this work

^{||} Present address: Leibniz Institut für Neurobiologie, 39118 Magdeburg, Germany

Two main forms of long-term potentiation (LTP)—a prominent model for the cellular mechanism of learning and memory—have been distinguished in the mammalian brain¹. One requires activation of postsynaptic NMDA (*N*-methyl *D*-aspartate) receptors, whereas the other, called mossy fibre LTP, has a principal presynaptic component. Mossy fibre LTP is expressed in hippocampal mossy fibre synapses^{1,2}, cerebellar parallel fibre synapses^{3,4} and corticothalamic synapses⁵, where it apparently operates by a mechanism that requires activation of protein kinase A. Thus, presynaptic substrates of protein kinase A are probably essential

in mediating this form of long-term synaptic plasticity. Studies of knockout mice have shown that the synaptic vesicle protein Rab3A is required for mossy fibre LTP⁶, but the protein kinase A substrates rabphilin, synapsin I and synapsin II are dispensable^{7,8}. Here we report that mossy fibre LTP in the hippocampus and the cerebellum is abolished in mice lacking RIM1 α , an active zone protein that binds to Rab3A and that is also a protein kinase A substrate^{9,10}. Our results indicate that the long-term increase in neurotransmitter release during mossy fibre LTP may be mediated by a unitary mechanism that involves the GTP-dependent interaction of Rab3A with RIM1 α at the interface of synaptic vesicles and the active zone.

The RIM family of active zone proteins may function as protein scaffolds that help to regulate vesicle exocytosis during short-term synaptic plasticity¹¹. To determine whether RIMs are essential at synapses expressing mossy fibre LTP (mflTP), we examined short- and long-term synaptic plasticity at hippocampal mossy fibre synapses and cerebellar parallel fibre/Purkinje cell synapses in mice lacking RIM1 α . Similar to hippocampal CA1 synapses¹¹, mossy fibre synapses are functional in RIM1 α knockout mice, and comparable synaptic responses could be elicited in wild-type and mutant synapses. We examined two forms of presynaptic short-term plasticity: paired-pulse facilitation (PPF)¹² and frequency facilitation¹³. To ensure the most accurate postsynaptic measure of presynaptic changes in transmitter release, we recorded isolated NMDA receptor (NMDAR)-mediated excitatory postsynaptic currents (EPSCs) at positive holding potentials (see Methods).

No differences in PPF monitored at three interstimulus intervals

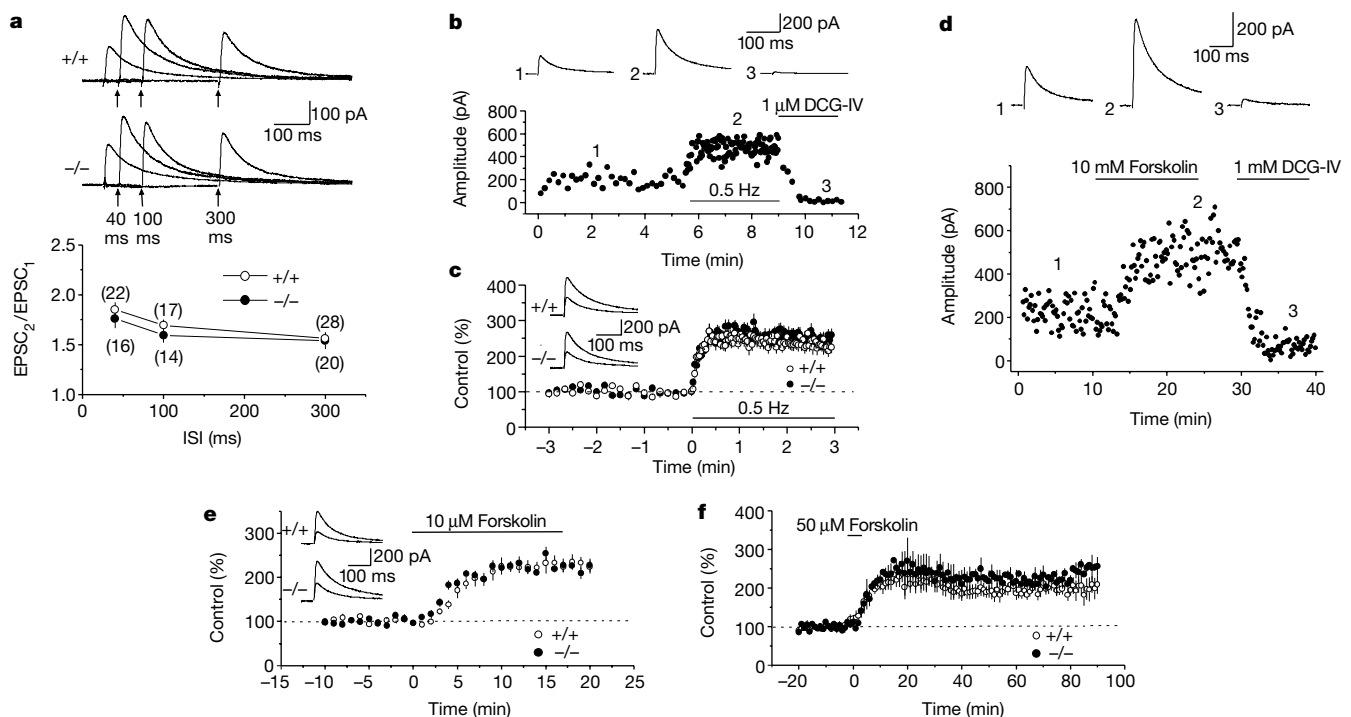


Figure 1 Short-term plasticity and forskolin-induced potentiation at hippocampal mossy fibre synapses in RIM1 α knockout and wild-type mice. Synaptic transmission was assessed by monitoring NMDAR-mediated EPSCs. **a**, Paired-pulse facilitation measured at three interstimulus intervals (ISIs). Top, typical EPSCs from a wild-type (+/+) and a RIM1 α knockout (-/-) mouse; paired-pulse responses are shown after subtraction of the first EPSC. Bottom, magnitude of paired-pulse facilitation (mean $\pm$ s.e.m.) is summarized for wild-type ($n = 6$) and knockout ($n = 7$) mice. Number of cells analysed is indicated. In all figures, traces are the average of 10–20 consecutive synaptic responses. **b, c**, Representative example (**b**) and summary (**c**) of frequency facilitation in wild-type

(14 cells/5 mice) and RIM1 α knockout (15/4) mice. Horizontal bar indicates stimulation rate change from 0.1 to 0.5 Hz. Traces in **b** were taken at the times indicated. Traces in **c** were taken during 0.1-Hz (thin trace) and 0.5-Hz (thick trace) stimulation. **d**, Example of synaptic enhancement elicited by forskolin. Traces were taken at the times indicated. **e**, Summary of forskolin's effects in wild-type (6 cells/3 mice) and RIM1 α knockout (6/3) mice. Traces were taken before and after forskolin application. **f**, Summary of effects of brief forskolin application in wild-type (3 cells/2 mice) and RIM1 α knockout (3/2) mice.

(40, 100 and 300 ms) were observed between the RIM1 α knockout and wild-type mice (Fig. 1a). Frequency facilitation, elicited by increasing the stimulation rate from 0.1 to 0.5 Hz (Fig. 1b, c), also was not affected in the RIM1 α knockout mice (wild type: $234 \pm 17\%$ of control baseline, 15 cells/4 mice; knockout: $251 \pm 16\%$, 14 cells/4 mice; $P > 0.5$). Thus, the loss of RIM1 α and the 60% decrease of Munc13-1 (ref. 11) does not produce an alteration in baseline synaptic transmission or short-term plasticity in mossy fibre synapses.

Increases in cyclic AMP, for example by forskolin application, cause a massive enhancement of glutamate release at mossy fibre synapses^{14,15}. The main mechanism of action of cAMP is probably the stimulation of protein kinase A (PKA), which has been implicated in mLT^{1,14,16}, but other, more direct effects of cAMP on transmitter release are also likely to operate¹⁵. As RIM1 α is an efficient substrate of PKA *in vivo* (G. Lonart, S.S. and T.C.S., unpublished data) and binds cAMP-GEFII (ref. 17), we examined whether the cAMP-mediated increase in synaptic transmission requires RIM1 α by testing the effects of forskolin (10 μ M for 15 min) in wild-type and RIM1 α -deficient mossy fibre synapses (Fig. 1d).

To minimize the contribution of voltage-dependent conductance and recurrent excitation to our measurements, we again monitored NMDAR-mediated EPSCs. Under these recording conditions, the effects of forskolin were significantly less than those observed with mossy fibre field potentials⁶. However, no differences were detected in the magnitude of the forskolin-induced enhancement in RIM1 α knockout mice (Fig. 1e; wild type: $216 \pm 16\%$ of control baseline, 7 cells/3 mice; knockout: $226 \pm 9\%$, 7 cells/3 mice; $P > 0.5$). This result is similar to that obtained in Rab3A, PKA and adenylyate

cyclase type I knockout animals, in which the forskolin response is also unaltered^{6,18,19}.

The potentiation of mossy fibre synaptic transmission induced by forskolin is long lasting, as its effect is maintained on washout¹⁴. Thus, it is possible that the persistent activation of PKA owing to application of forskolin may have obscured a deficit in the RIM1 α knockout mice. We addressed this possibility by measuring the effects of a brief application of forskolin (50 μ M for 3 min; Fig. 1f). This long-lasting forskolin-induced potentiation, measured 90 min after forskolin application, was similar in wild-type and RIM1 α knockout mice (wild type: $198 \pm 19\%$ of control baseline, 3 cells/2 mice; knockout: $225 \pm 16\%$, 3 cells/2 mice), indicating that the increase in transmitter release caused by raising cAMP concentrations is not affected significantly by the absence of RIM1 α .

We next tested whether mLT¹ was normal in RIM1 α knockout mice (Fig. 2a, b). In wild-type animals, robust mLT¹ was induced by a single tetanus (25 Hz for 5 s in the presence of 50 μ M AP5 (D,-)-2-amino-5-phosphonopivalic acid); $170 \pm 13\%$ of baseline measured 45–60 min after the tetanus, 12 slices, 5 mice). By contrast, the same induction protocol failed to elicit LTP in RIM1 α knockout mice ($106 \pm 6\%$ of baseline, 14 slices, 6 mice; $P < 0.002$). mLT¹ is thought to be triggered by a rise of presynaptic Ca^{2+} (ref. 20), although a postsynaptic Ca^{2+} signal has also been suggested²¹. As the absence of RIM1 α might influence the magnitude of the rise in Ca^{2+} or the way that Ca^{2+} is handled after repetitive stimulation, we determined whether a manipulation that greatly increases tetanus-induced Ca^{2+} influx could rescue mLT¹ in the RIM1 α knockout mice. Specifically, we doubled the extracellular Ca^{2+} concentration (from 2.5 to 5.0 mM), and delivered four tetani (25 Hz for 5 s with 20-s intervals) instead of one. Increasing extracellular Ca^{2+} elicited a clear enhancement of mossy fibre synaptic responses as expected (Fig. 2c); however, LTP was still abolished in the RIM1 α knockout mice (wild type: $219 \pm 24\%$ of baseline measured 50–65 min after induction, 6 slices, 3 mice; knockout: $103 \pm 3\%$ of baseline; 5 slices, 3 mice; $P < 0.005$).

The deletion of RIM1 α caused about a 60% decrease in the

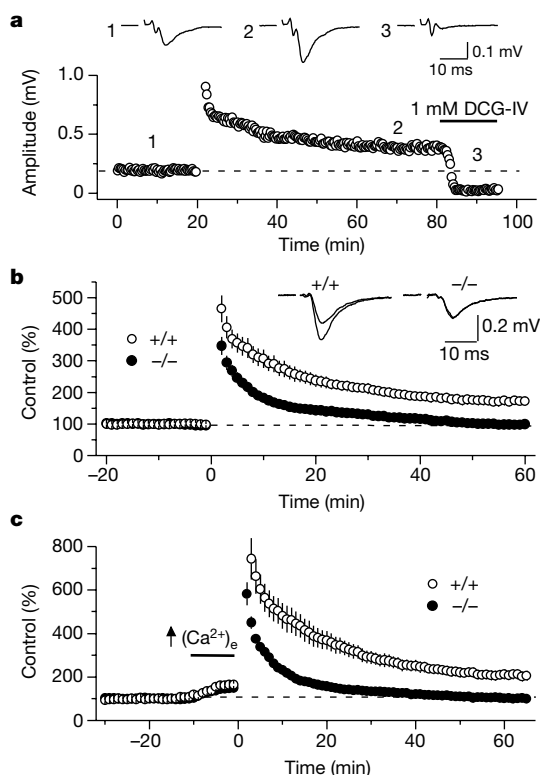


Figure 2 Mossy fibre LTP is abolished in RIM1 α knockout mice. **a**, **b**, Example (**a**) and summary (**b**) of mLT¹ in wild-type (12 slices/5 mice) and knockout (14/6) mice. Traces in **a** were taken at the times indicated. Traces in **b** were taken before (thin trace) and 55–60 min after (thick trace) tetanus. **c**, Summary of LTP induced by four trains (5 s at 25 Hz, every 20 s) after increasing extracellular Ca^{2+} from 2.5 mM to 4.0 mM in wild-type (6 slices/3 mice) and RIM1 α knockout (5/3) mice.

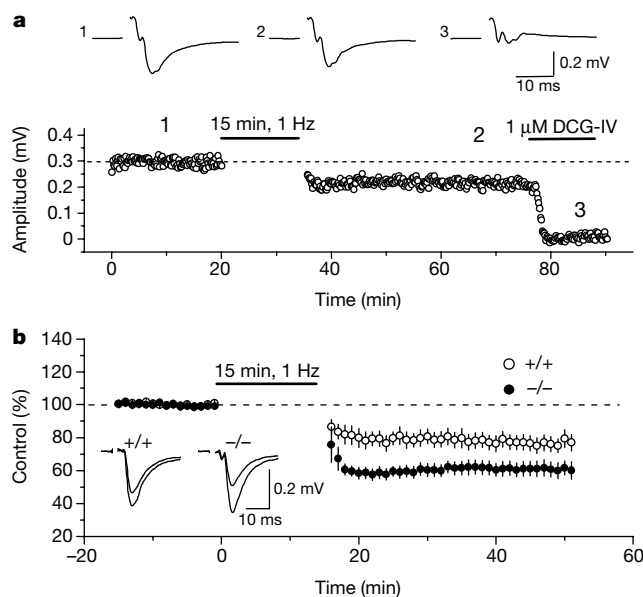


Figure 3 Mossy fibre LTD is enhanced in RIM1 α knockout mice. Example (**a**) and summary (**b**) of mossy fibre LTD in wild-type (10 slices/4 mice) and RIM1 α knockout (11/5) mice. Traces in **a** were taken at the times indicated. Traces in **b** were taken before (thin trace) and after (thick trace) the tetanus (900 stimuli at 1 Hz) used to induce LTD.

concentration of total brain Munc13-1 (ref. 11), presumably because the binding of Munc13-1 to RIM1 α stabilizes it and prevents its degradation^{11,22}; thus, it is possible that the lack of mflTP in the RIM1 α knockout mice might be due to this decrease in Munc13-1. We therefore examined mflTP in Munc13-1 heterozygotes that express roughly 50% of the normal concentration of Munc13-1 (ref. 11). Slices prepared from Munc13-1 heterozygotes displayed normal mflTP (Munc13-1^{+/-}: $170 \pm 13\%$ of baseline, 5 slices, 3 mice; wild-type littermates: $175 \pm 11\%$ of baseline, 4 slices, 2 mice; data not shown), indicating that the decrease in Munc13-1 concentration in the RIM1 α knockout mice does not cause their inability to express mflTP.

As previous work has suggested that Rab3A also may be essential for long-term depression (LTD) at mossy fibre synapses²³, we examined mossy fibre LTD in RIM1 α knockout mice. We found that not only was mossy fibre LTD present in RIM1 α knockout mice, but its magnitude was significantly larger than in wild-type animals (Fig. 3; wild type: $77 \pm 5\%$ of baseline, 10 slices, 4 mice; knockout: $63 \pm 4\%$ of baseline, 11 slices, 5 mice; $P < 0.05$). One explanation for this observation is that the LTD induction protocol, by raising the concentration of Ca²⁺ in the mossy fibre terminal, also induces a degree of LTP that reduces the apparent magnitude of LTD in wild-type but not knockout animals.

As these results indicated that Rab3A and RIM1 α may have distinct actions in LTD at mossy fibre synapses, we re-examined mossy fibre LTD in the Rab3A knockout mice. Unexpectedly we found that LTD was present in the Rab3A knockout mice (wild-type: $76 \pm 2\%$ of baseline, 4 slices, 2 mice; Rab3A knockout: $70 \pm 5\%$ of baseline, 4 slices, 2 mice), indicating that, like RIM1 α , Rab3A is not essential for LTD. (Changes in experimental conditions (such as animal age) are likely to be primarily responsible for the different results on mossy fibre LTD in the Rab3A knockout mice reported previously²³.)

Our results indicate that RIM1 α may have a mandatory role in the long-term enhancement of neurotransmitter release from mossy fibre terminals. Is this function of RIM1 α limited to mossy fibre synapses that are notable for unusual synaptic and morphological properties², or is it generally applicable to PKA-dependent pre-synaptic LTP at other synapses? To address this question, we examined LTP at cerebellar parallel fibre/Purkinje cell synapses that are morphologically and functionally distinct from mossy fibre synapses. Like mossy fibre synapses, we readily elicited apparently normal EPSCs on stimulation of parallel fibre synapses in the RIM1 α knockout mice, suggesting that they have a functional synaptic apparatus. Again, no significant change in short-term synaptic plasticity was observed as there was no detectable difference in PPF tested at four interstimulus intervals (Fig. 4a). To examine parallel fibre LTP, two independent sets of parallel fibres were activated: LTP was induced in one pathway by tetanic stimulation (100 stimuli at 10 Hz); and the other pathway served as a control (Fig. 4b). In wild-type animals, clear LTP was elicited ($158 \pm 10\%$ of baseline measured at 25 to 30 min after tetanus, 13 cells/4 mice), whereas LTP was almost completely abolished in RIM1 α knockout mice ($112 \pm 15\%$ of baseline, 15 cells/5 mice; $P < 0.001$; Fig. 4c).

The synaptic vesicle protein Rab3A is essential for inducing LTP in mossy fibre synapses⁶, but the mechanism by which Rab3A performs this role is unclear, in particular because Rab3A itself is not a PKA substrate. We have shown here that RIM1 α , a presynaptic PKA substrate that is a putative GTP-dependent effector for Rab3A, is not required for basal mossy fibre synaptic transmission or short-term presynaptic plasticity, but is essential for mflTP. This phenotype is unlikely to be caused by structural abnormalities, because light and electron microscopy both showed that mossy fibre terminals have an overall normal structure and organization in RIM1 α knockout animals (data not shown). We also found similar effects of the RIM1 α deletion at parallel fibre synapses in the

cerebellum, providing further evidence that LTP at mossy fibre and parallel fibre synapses have the same molecular basis. Our results localize the expression of mflTP to the interface between synaptic vesicles and the active zone, as a pair of interacting proteins, the vesicle protein Rab3A and the active zone protein RIM1 α , are required. Our results, however, do not rule out the participation of additional proteins in this presynaptic form of LTP.

Because of the large changes in release probability and short-term synaptic plasticity observed at CA1 excitatory synapses in RIM1 α knockout mice¹¹, it is surprising that CA3 mossy fibre synapses and parallel fibre synapses in these mice show only impaired LTP. A similar selectivity was observed previously at mossy fibre synapses in Rab3A knockout animals⁶. An explanation for these differences between synapses is suggested by the fact that PKA-dependent mflTP can only be elicited in a subset of synapses, which indicates that the molecular machinery available for the regulation of neurotransmitter release is different between mflTP-competent

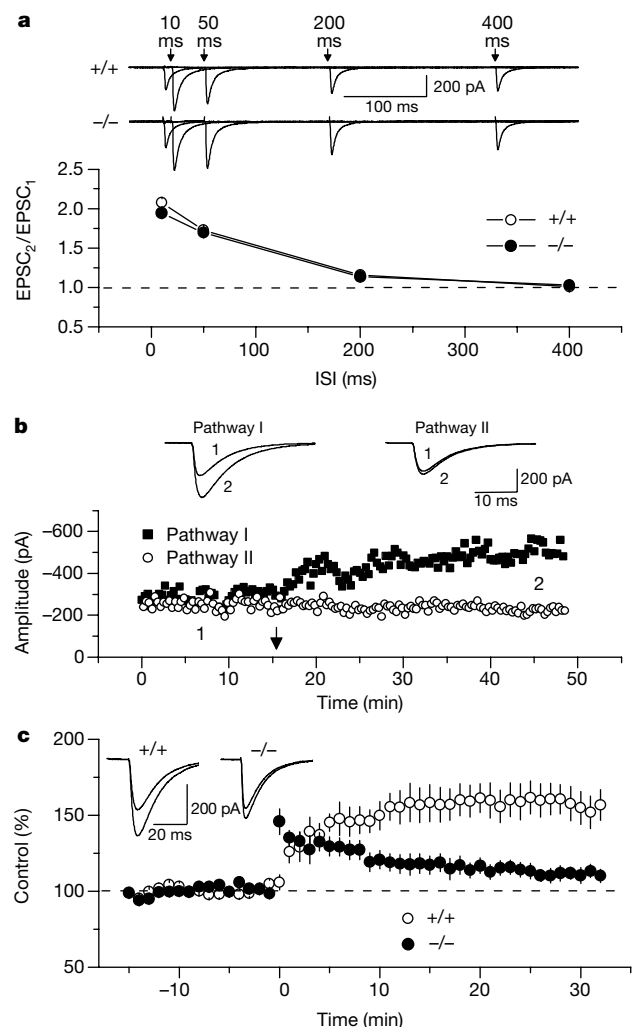


Figure 4 Paired-pulse facilitation and LTP at cerebellar parallel fibre synapses in wild-type and RIM1 α knockout mice. **a**, Summary of paired-pulse facilitation measured at four interstimulus intervals (ISIs) in wild-type (15 cells/4 mice) and RIM1 α knockout (16/4) mice. Traces are shown after subtraction of the first EPSC. **b**, **c**, Example **(b)** and summary **(c)** of LTP in wild-type (13 cells/4 mice) and RIM1 α knockout (15/5) mice. Two pathways were monitored, and after the responses stabilized, one pathway was tetanized (100 stimuli at 10 Hz; arrow in **b**; time 0 in **c**). Traces in **b** were taken at the times indicated. Traces in **c** were taken before (thin trace) and 30 min after (thick trace) tetanus.

(such as mossy fibre synapses and parallel fibre synapses) and mLTIP-incompetent synapses (such as CA1 synapses). The normal synaptic enhancement elicited by forskolin in the RIM1 α knockout mice may also seem puzzling. PKA-independent actions of cAMP, however, may account for the action of forskolin in the RIM1 α knockout mice^{15,24–26}.

Our results have provided further elucidation of the molecular pathway for a principal form of LTP in the mammalian brain. They indicate that RIM1 α may be an effector of Rab3A at synapses expressing mLTIP, in contrast to the more abundant putative Rab3A effector, rabphilin⁷. The results in this and the accompanying paper¹¹ indicate that there is significant molecular heterogeneity between synapses and that RIMs, because of their role as ubiquitous protein scaffolds in presynaptic terminals, probably have a key role at all synapses in coordinating the active zone and synaptic vesicle proteins that are ultimately responsible for both the short-term and long-term regulation of neurotransmitter release. □

Methods

Mossy fibre to CA3 pyramidal cell synapses

Standard procedures²⁰ were used for preparing hippocampal slices (400- μ m thick) from mice aged 3–6 weeks. The composition of the external recording solution was (in mM): 119 NaCl, 2.5 KCl, 1.3 MgSO₄, 2.5 CaCl₂, 1.0 NaH₂PO₄, 26.2 NaHCO₃ and 10 D-glucose (saturated with 95% O₂ and 5% CO₂). Mossy fibre synaptic responses were monitored at room temperature either by extracellular field potentials (fEPSPs) from the stratum lucidum in the CA3 area or, by recording synaptic currents from CA3 pyramidal cells in whole-cell configuration²⁰.

We performed extracellular recordings with a patch pipette filled with 1 M NaCl. For whole-cell voltage-clamp recording, the internal solution contained (in mM): 123 caesium gluconate, 10 CsCl, 8 NaCl, 1 CaCl₂, 10 EGTA, 10 HEPES, 10 glucose (pH 7.2; 280 mOsm). Series resistance (10–20 M Ω) was monitored throughout each experiment with an 80-ms pulse of –4 mV amplitude every 20 s; those cells in which series resistance changed more than 15% were systematically excluded from this study. We stimulated mossy fibres by placing a whole-cell recording pipette filled with extracellular solution in the dentate gyrus or stratum lucidum of the CA3 cell body layer. Baseline stimulation rate was 0.1 Hz for all experiments.

To avoid recurrent contamination from associational/commissural fibres or di-synaptic activation of pyramidal cells, for some experiments the mossy fibre NMDAR-mediated EPSC (NMDAR EPSC) was monitored in the presence of the AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid)/kainate receptor antagonist NBQX (10 μ M). In these experiments, CA3 pyramidal cells were held at +30 to +40 mV and 100 μ M picrotoxin was also included in the bath to block GABA (γ -amino butyric acid)-mediated inhibitory inputs. In particular, we recorded NMDAR EPSCs whenever the enhancement of synaptic transmission was enough to fire action potentials from the postsynaptic cell; such firing makes the fEPSP a potentially poor measure of quantitative changes in the strength of mossy fibre synaptic transmission.

In all experiments, to confirm that fEPSPs and NMDAR EPSCs were caused by activation of mossy fibres, 1 μ M DCG-IV, a metabotropic glutamate receptor agonist that selectively blocks mossy fibre synaptic transmission²⁷, was applied at the end of the experiment and data were included only if the responses were reduced by more than 90%. Furthermore, we examined the occurrence of large PPF and frequency facilitation, both of which are prominent at mossy fibre synapses but not at associational/commissural synapses²⁸. For all experiments, the amplitude of mossy fibre responses was measured after subtraction of the response that remained in the presence of DCG-IV.

Activation of presynaptic kainate receptors contributes to frequency facilitation at mossy fibre synapses²⁹, an effect that might mask differences between wild-type and mutant animals. To test whether presynaptic kainate receptors contributed to the short-term plasticity that we measured, we compared in the same CA3 pyramidal cell, frequency facilitation of NMDAR EPSCs in the presence of 10 μ M NBQX (the concentration used in the preceding experiments) and then 100 μ M NBQX. There was no significant difference in frequency facilitation under these two conditions ($n = 4$ cells; data not shown) indicating that 10 μ M NBQX was sufficient to block the presynaptic effects of kainate receptor activation at mossy fibre synapses.

Parallel fibre to Purkinje cell synapses

Parasagittal slices (400- μ m thick) from cerebellar vermis were prepared as described³. Whole-cell recordings from Purkinje cells were performed under visual control with an upright Olympus microscope (40 $\times$ water-immersion objective) and infrared differential interference contrast videomicroscopy. Cells were voltage clamped at –70 mV and recorded at room temperature. The composition of the recording pipette solution was the same as used for CA3 pyramidal cells; series resistance was typically 8–12 M Ω throughout the experiment. Two independent sets of parallel fibres were stimulated extracellularly with monopolar stimulating electrodes consisting of glass patch pipettes filled with bath solution; these stimulating pipettes were placed 200–400- μ m apart in the middle third of the molecular layer, roughly equidistant from the recorded cell. The baseline parallel-fibre-

evoked EPSC amplitude was kept less than 300 pA to avoid sodium spikes that escaped voltage clamp, in particular, during and after tetanization.

Data analysis

All summary data are expressed as the mean $\pm$ s.e.m. as a percentage of the baseline. We used Student's t -test to determine whether there was a significant difference in the means between the results from wild-type and mutant mice. Mice were shipped to Stanford Univ. (Palo Alto, California) unidentified, and data were acquired and analysed in a blind fashion; the genotype was confirmed afterwards by PCR using tail DNA.

Received 26 September; accepted 25 October 2001.

- Nicoll, R. A. & Malenka, R. C. Contrasting properties of two forms of long-term potentiation in the hippocampus. *Nature* **377**, 115–118 (1995).
- Henze, D. A., Urban, N. N. & Barrionuevo, G. The multifarious hippocampal mossy fiber pathway: a review. *Neuroscience* **98**, 407–427 (2000).
- Salin, P. A., Malenka, R. C. & Nicoll, R. A. Cyclic AMP mediates a presynaptic form of LTP at cerebellar parallel fiber synapses. *Neuron* **16**, 797–803 (1996).
- Linden, D. J. & Ahn, S. Activation of presynaptic cAMP-dependent protein kinase is required for induction of cerebellar long-term potentiation. *J. Neurosci.* **19**, 10221–10227 (1999).
- Castro-Alamancos, M. A. & Calcagno, M. E. Presynaptic long-term potentiation in corticothalamic synapses. *J. Neurosci.* **19**, 9090–9097 (1999).
- Castillo, P. E. et al. Rab3A is essential for mossy fibre long-term potentiation in the hippocampus. *Nature* **388**, 590–593 (1997).
- Schlüter, O. M. et al. Rabphilin knock-out mice reveal that rabphilin is not required for rab3 function in regulating neurotransmitter release. *J. Neurosci.* **19**, 5834–5846 (1999).
- Spillane, D. M., Rosahl, T. W., Südhof, T. C. & Malenka, R. C. Long-term potentiation in mice lacking synapsins. *Neuropharmacology* **34**, 1573–1579 (1995).
- Wang, Y., Okamoto, M., Schmitz, F., Hofmann, K. & Südhof, T. C. Rim is a putative Rab3 effector in regulating synaptic-vesicle fusion. *Nature* **388**, 593–598 (1997).
- Wang, Y., Sugita, S. & Südhof, T. C. The RIM/NIM family of neuronal C2 domain proteins. Interactions with Rab3 and a new class of Src homology 3 domain proteins. *J. Biol. Chem.* **275**, 20033–20044 (2000).
- Schoch, S. et al. RIM1 α forms a protein scaffold for regulating neurotransmitter release at the active zone. *Nature* **415**, 321–326 (2002).
- Zucker, R. S. Short-term synaptic plasticity. *Annu. Rev. Neurosci.* **12**, 13–31 (1989).
- Regehr, W. G., Delaney, K. R. & Tank, D. W. The role of presynaptic calcium in short-term enhancement at the hippocampal mossy fiber synapse. *J. Neurosci.* **14**, 523–537 (1994).
- Weisskopf, M. G., Castillo, P. E., Zalutsky, R. A. & Nicoll, R. A. Mediation of hippocampal mossy fiber long-term potentiation by cyclic AMP. *Science* **265**, 1878–1882 (1994).
- Lonart, G., Janz, R., Johnson, K. M. & Südhof, T. C. Mechanism of action of rab3A in mossy fiber LTP. *Neuron* **21**, 1141–1150 (1998).
- Huang, Y. Y., Li, X. C. & Kandel, E. R. cAMP contributes to mossy fiber LTP by initiating both a covalently mediated early phase and macromolecular synthesis-dependent late phase. *Cell* **79**, 69–79 (1994).
- Ozaki, N. et al. cAMP-GEFII is a direct target of cAMP in regulated exocytosis. *Nature Cell Biol.* **2**, 805–811 (2000).
- Villacres, E. C., Wong, S. T., Chavkin, C. & Storm, D. R. Type I adenylyl cyclase mutant mice have impaired mossy fiber long-term potentiation. *J. Neurosci.* **18**, 3186–3194 (1998).
- Huang, Y. Y. et al. A genetic test of the effects of mutations in PKA on mossy fiber LTP and its relation to spatial and contextual learning. *Cell* **83**, 1211–1222 (1995).
- Castillo, P. E., Weisskopf, M. G. & Nicoll, R. A. The role of Ca²⁺ channels in hippocampal mossy fiber synaptic transmission and long-term potentiation. *Neuron* **12**, 261–269 (1994).
- Yeckel, M. F., Kapur, A. & Johnston, D. Multiple forms of LTP in hippocampal CA3 neurons use a common postsynaptic mechanism. *Nature Neurosci.* **2**, 625–633 (1999).
- Betz, A. et al. Functional interaction of the active zone proteins Munc13-1 and RIM1 in synaptic vesicle priming. *Neuron* **30**, 183–196 (2001).
- Tzounopoulos, T., Janz, R., Südhof, T. C., Nicoll, R. A. & Malenka, R. C. A role for cAMP in long-term depression at hippocampal mossy fiber synapses. *Neuron* **21**, 837–845 (1998).
- Beaumont, V. & Zucker, R. S. Enhancement of synaptic transmission by cyclic AMP modulation of presynaptic Ih channels. *Nature Neurosci.* **3**, 133–141 (2000).
- Kawasaki, H. et al. A family of cAMP-binding proteins that directly activate Rap1. *Science* **282**, 2275–2279 (1998).
- de Rooij, J. et al. Epac is a Rap1 guanine-nucleotide-exchange factor directly activated by cyclic AMP. *Nature* **396**, 474–477 (1998).
- Kamiya, H., Shinozaki, H. & Yamamoto, C. Activation of metabotropic glutamate receptor type 2/3 suppresses transmission at rat hippocampal mossy fiber synapses. *J. Physiol. (Lond.)* **493**, 447–455 (1996).
- Salin, P. A., Scanziani, M., Malenka, R. C. & Nicoll, R. A. Distinct short-term plasticity at two excitatory synapses in the hippocampus. *Proc. Natl Acad. Sci. USA* **93**, 13304–13309 (1996).
- Schmitz, D., Mellor, J. & Nicoll, R. A. Presynaptic kainate receptor mediation of frequency facilitation at hippocampal mossy fiber synapses. *Science* **291**, 1972–1976 (2001).

Acknowledgements

We thank H. Riedesel for help with the RIM1 α knockout mice. This study was supported by grants from the NIH (R.C.M.).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.C.M. (e-mail: malenka@stanford.edu).

Replication-incompetent adenoviral vaccine vector elicits effective anti-immunodeficiency-virus immunity

John W. Shiver*, Tong-Ming Fu*, Ling Chen*, Danilo R. Casimiro*, Mary-Ellen Davies*, Robert K. Evans*, Zhi-Qiang Zhang*, Adam J. Simon*, Wendy L. Triglona*, Sheri A. Dubey*, Lingyi Huang*, Virginia A. Harris*, Ronnie S. Long*, Xiaoping Liang*, Larry Handt*, William A. Schleif*, Lan Zhu*, Daniel C. Freed*, Natasha V. Persaud*, Liming Guan*, Kara S. Punt*, Aimin Tang*, Minchun Chen*, Keith A. Wilson*, Kelly B. Collins*, Gwendolyn J. Heidecker*, V. Rose Fernandez*, Helen C. Perry*, Joseph G. Joyce*, Karen M. Grimm*, James C. Cook*, Paul M. Keller*, Denise S. Kresock*, Henryk Mach*, Robert D. Troutman*, Lynne A. Isopi*, Donna M. Williams*, Zheng Xu*, Kathryn E. Bohannon*, David B. Volkin*, David C. Montefiori†, Ayako Miura‡, Georgia R. Krivulka‡, Michelle A. Lifton‡, Marcelo J. Kuroda‡, Jörn E. Schmitz‡, Norman L. Letvin‡, Michael J. Caulfield*, Andrew J. Bett*, Rima Youil*, David C. Kaslow* & Emilio A. Emini*

* Merck Research Laboratories, West Point, Pennsylvania 19486, USA

† Center for AIDS Research, Duke University Medical Center, Durham, North Carolina 27710, USA

‡ Division of Viral Pathogenesis, Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts 02215, USA

Recent studies of human immunodeficiency virus type 1 (HIV-1) infection in humans and of simian immunodeficiency virus (SIV) in rhesus monkeys have shown that resolution of the acute viral infection and control of the subsequent persistent infection are mediated by the antiviral cellular immune response^{1–11}. We comparatively assessed several vaccine vector delivery systems—three formulations of a plasmid DNA vector, the modified vaccinia Ankara (MVA) virus, and a replication incompetent adenovirus type 5 (Ad5) vector—expressing the SIV gag protein for their ability to elicit such immune responses in monkeys. The vaccines were tested either as a single modality or in combined modality regimens. Here we show that the most effective responses were

elicited by a replication-incompetent Ad5 vector, used either alone or as a booster inoculation after priming with a DNA vector. After challenge with a pathogenic HIV–SIV hybrid virus (SHIV), the animals immunized with Ad5 vector exhibited the most pronounced attenuation of the virus infection. The replication-defective adenovirus is a promising vaccine vector for development of an HIV-1 vaccine.

Two independently performed immunization and viral challenge studies (studies A and B) were conducted in rhesus macaque monkeys (*Macaca mulatta*). The immunization groups and experimental design for each study are summarized in Table 1. In study A, 21 rhesus monkeys were divided into six groups, including an unimmunized control cohort of six animals and five immunized groups of three monkeys each. Study B used 14 monkeys, three each in the immunized groups and eight in the unimmunized control group. Each of the test vectors expressed the identical SIVmac239 gag gene that had been codon optimized for expression in mammalian cells. The plasmid DNA vector was formulated in either phosphate-buffered saline (PBS) solution, a solution containing a nonionic blocked copolymer adjuvant (CRL1005)¹², or a solution containing monophosphoryl lipid A adsorbed onto aluminium phosphate (MPL/alum). All immunized animals were genotyped for the MHC class I *Mamu-A*01* allele and, with the exception of one monkey in the DNA/PBS group, all were positive for this allele. Two of the six and four of the eight control animals in studies A and B, respectively, were also positive for *Mamu-A*01*. The presence of this MHC allele allowed us to analyse the responses of T cells positive for the CD8 antigen (CD8⁺) to the test vaccines; we used a tetramer reagent designed to detect T-cell receptors capable of binding an immunodominant SIV gag epitope, p11CM (residues 181–189), presented by the *Mamu-A*01* MHC protein^{13,14}.

All the vaccines were delivered by intramuscular inoculation. In the first study, the two viral vector vaccines were administered at weeks 0 and 6 followed by a booster dose at week 32; the preparations of DNA plasmid vector vaccine were delivered at weeks 0, 4 and 8 followed by a booster at week 25. In the second study, the DNA vector priming inoculations were also administered at weeks 0, 4 and 8. The viral vector boost inoculations were then given at week 32. As noted in Table 1, four of the control animals in study B (both positive and negative for *Mamu A*01*) were immunized with

Table 1 Immunization and challenge schedule

Group no.	Monkey no.	Priming immunization and schedule	Boost immunization and schedule	Challenge*
Study A				
1	115Q, V388, 127T†, 068T†, V446†, V391†	None	NA	NA
2	97X011, T271, T282†	SIV gag DNA, 5 mg, weeks 0, 4 and 8	SIV gag DNA, 5 mg, week 25	12
3	S202, 047R, 079R	SIV gag DNA with CRL1005, 5 mg, weeks 0, 4 and 8	SIV gag DNA with CRL1005, 5 mg, week 25	12
4	058R, 088R, 108R	SIV gag DNA with MPL/Alum, 5 mg, weeks 0, 4 and 8	SIV gag DNA with MPL/Alum, 5 mg, week 25	12
5	122G, 126G, 15G	MVA–SIV gag, 10 ⁹ p.f.u., weeks 0 and 6	MVA–SIV gag, 10 ⁹ p.f.u., week 32	12
6	54G, 56G, 82F	Ad5–SIV gag, 10 ¹¹ particles, weeks 0 and 6	Ad5–SIV gag, 10 ¹¹ particles, week 32	12
Study B				
7	040F‡, 97N114‡, 89Q, AW35, 99D142‡, 99D158‡, 99D221†, 99D149†	None	NA	NA
8	98D273, 98D318, 98D359	SIV gag DNA with CRL1005, 5 mg, weeks 0, 4 and 8	MVA–SIV gag, 10 ⁹ p.f.u., week 32	6
9	98C040, 98C081, 98C084	SIV gag DNA with CRL1005, 5 mg, weeks 0, 4 and 8	Ad5–SIV gag, 10 ¹¹ particles, week 32	6

See text for vaccine descriptions. The boost immunization in study A was conducted simultaneously for all groups in the study. NA, not applicable; p.f.u., plaque-forming units.

* Time of challenge is indicated as weeks after boost immunization. The animals in study A and study B were challenged separately. However, all animals within each study were challenged simultaneously.

† These monkeys did not express the *Mamu-A*01* MHC allele.

‡ These monkeys were mock immunized with 10¹¹ particles of empty Ad5 vector at week 32.

an 'empty' Ad5 vector (10^{11} particles) that did not contain the SIV *gag* gene as a control for possible nonspecific immune effects on viral challenge.

The levels of circulating p11CM-specific CD8⁺ T cells for the *Mamu-A*01*-expressing monkeys were monitored by the tetramer assay. In general, all of the monkeys developed p11CM-specific cellular immune responses after the initial immunization series (Fig. 1). These responses in study A were highest in the Ad5-vaccinated monkeys, followed in order by the DNA/CRL1005, MVA, DNA/MPL/alum and DNA/PBS cohorts. In the Ad5-vaccinated monkeys, the numbers of p11CM-specific cells reached levels as high as 4.0% of the circulating CD8⁺ T-cell population before declining to 0.3–1.0% at the time of the booster inoculation. For all three animals, administration of the third dose of the Ad5 vector resulted in an additional increase of p11CM-specific cells, with levels maintained at about 1.0–2.0% of circulating CD8⁺ T cells at the time of virus challenge. Staining with p11CM tetramers was notably lower at challenge in the monkeys from the remaining immunized groups within this study. The greatest p11CM-specific responses were seen in study B, in those monkeys that received priming immunizations with DNA/CRL1005 followed by the Ad5-vector boost (Fig. 1). After the booster inoculation, these animals exhibited peak levels of p11CM-specific T cells of 20.0–30.0% of circulating CD8⁺ T cells, with levels ranging from 5.0% to 25.0% at challenge. The group that received the prime-boost combination of DNA/CRL1005 and MVA exhibited post-boost peaks in p11CM tetramer staining of 3.0–5.0%, with levels of less than 1.0% at virus challenge. As expected, staining was not detected in any monkey after immunization using tetramer reagents for *Mamu-A*01*-restricted determinants in the viral *env* (p41, residues 431–439) and *pol* (p68, residues 621–629) proteins (data not shown)¹⁵.

The functional nature of the tetramer-positive CD8⁺ T cells was

confirmed in all of the immunized monkeys by an ELISPOT assay designed to determine the number of antigen-specific T cells that secreted interferon- γ (IFN- γ) when stimulated with either the p11CM peptide or a pool of peptides derived from SIV *gag* (see Supplementary Information Table 1). The relative levels of IFN- γ -secreting cells in the animals correlated well with the tetramer-staining results. Intracellular IFN- γ -staining assays confirmed that these responses were primarily mediated by CD8⁺ T cells (data not shown). The functional nature of the responding T cells in the immunized monkeys was further confirmed by a bulk culture killing assay for cytotoxic T lymphocytes (CTLs) (data not shown).

At 12 weeks (study A) or 6 weeks (study B) after the final immunization, all monkeys were challenged by intravenous injection with 50 MID₅₀ (50% monkey infectious doses) of the pathogenic SHIV 89.6P¹⁶. The challenge of the control and immunized animals within the context of each of the two independent studies occurred concurrently. Most of the immunized *Mamu-A*01*-positive monkeys responded to the challenge with a marked increase in staining of p11CM-specific CD8⁺ tetramers, reaching 10.0–40.0% of total CD3⁺ CD8⁺ circulating T cells at about 3 weeks after challenge (see Supplementary Information Fig. 1). Although the post-challenge expansion seen in the cohort of DNA/CRL1005 prime followed by Ad5 boost was relatively modest, the animals in this group maintained substantial levels of tetramer staining-specific cells, of 10.0–30.0%. By contrast, the *Mamu-A*01*-positive monkeys in both unimmunized control groups developed much weaker p11CM-tetramer responses after challenge.

Counts of CD4⁺ T cells and levels of plasma virus were followed over time for all monkeys after challenge (Fig. 2). Each of the animals in both control groups exhibited acute CD4⁺ T-cell lymphopenia and peak viral loads of 10^8 – 10^9 viral RNA copies per ml plasma at about 3 weeks after challenge, consistent with reported

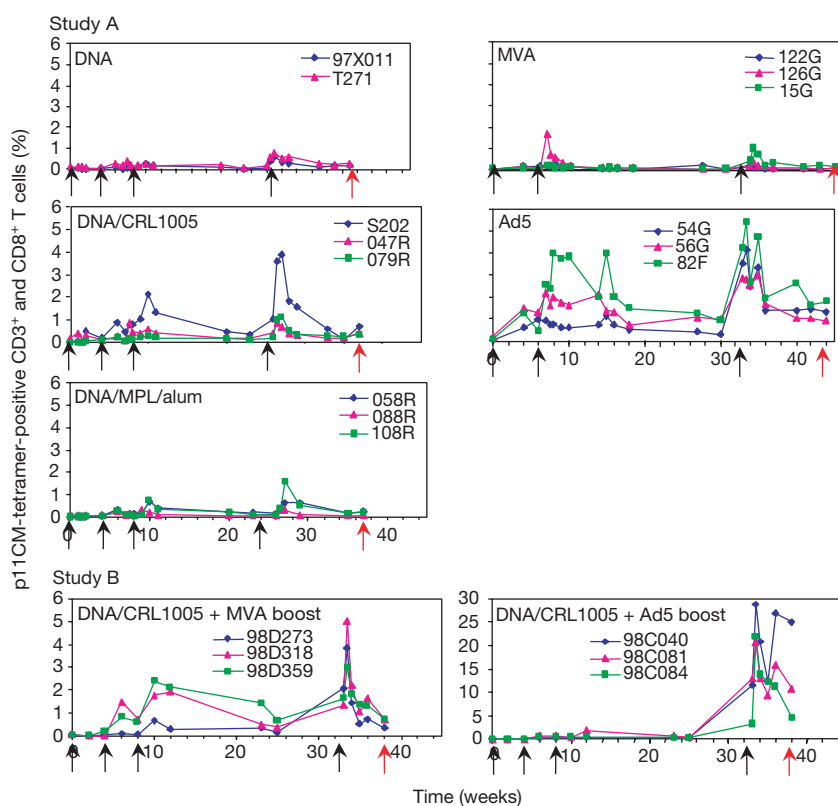


Figure 1 Quantification of CD8⁺ T-cell responses by tetramer analysis during immunization. The percentage of CD3⁺ CD8⁺ cells that bind tetramers are shown for each vaccine group. The black arrows indicate the time of each immunization; the red arrow

represents the time of SHIV challenge. Note that the y-axis scale in the right panel of study B is different from the remaining panels.

Study A

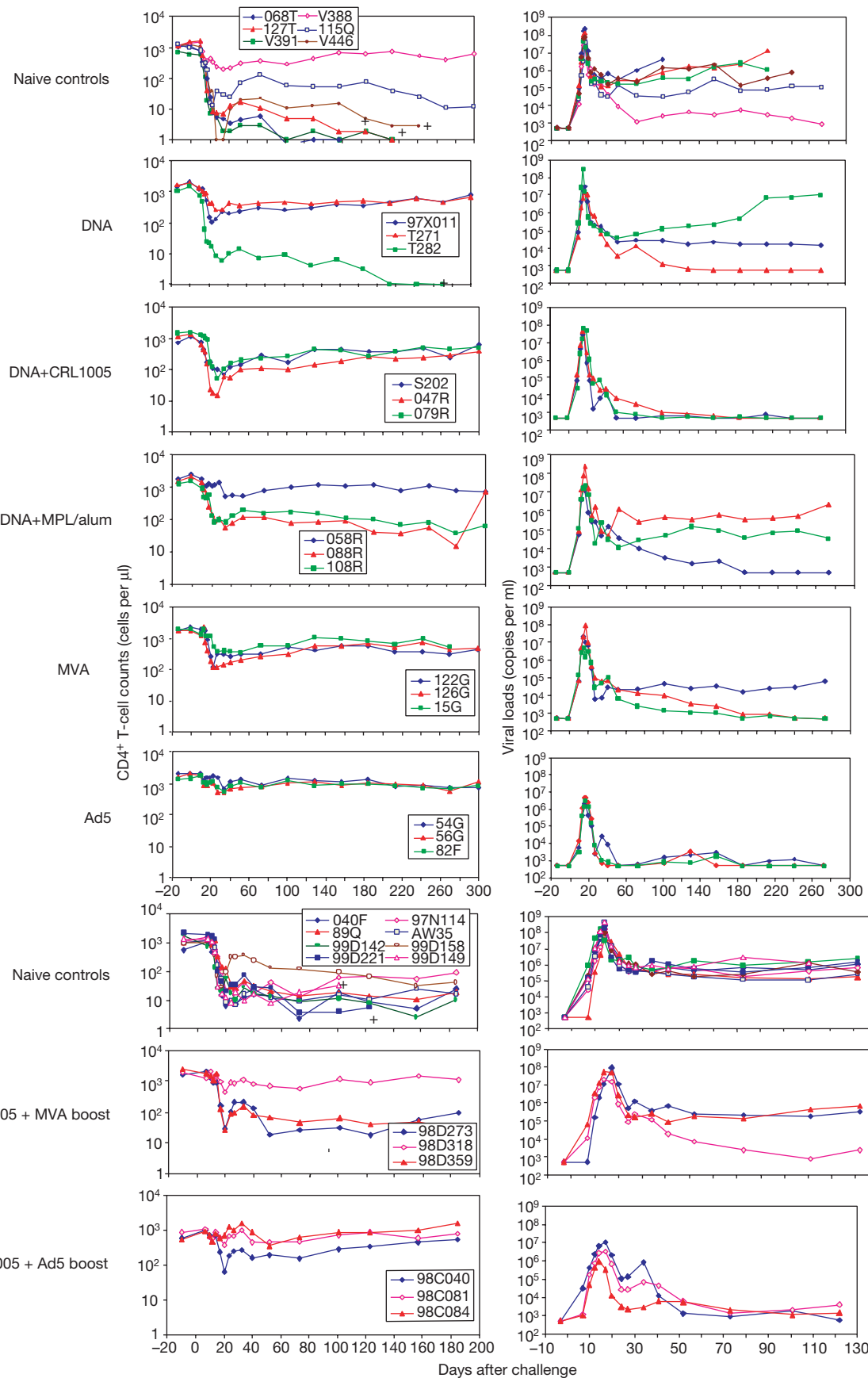


Figure 2 Post-challenge viraemia and CD4⁺ T-cell counts. For CD4⁺ cell counts, total lymphocyte numbers were multiplied by the per cent CD3⁺ CD4⁺ lymphocyte staining based on flow cytometry and reported as the number of cells per microlitre of whole blood.

Plasma viral levels were determined by a branched DNA-amplification assay with a detection limit of 500 viral RNA copies per ml (Bayer Diagnostics).

characteristics of infection with this virus in rhesus monkeys¹⁶. With the exception of one animal in study A, all of the control monkeys experienced dramatic loss of CD4⁺ T cells, with counts remaining below 100 cells μL^{-1} during the chronic phase of the infection. In these animals, plasma viral loads did not drop below 10^5 vRNA copies per ml. These observations were noted in both *Mamu-A*01*-positive and -negative animals, suggesting that the expression of this specific *Mamu-A*01* allele does not influence the course of infection after challenge with SHIV 89.6P. Control animals in study B that were immunized with an 'empty' Ad5 vector did not exhibit any differences in plasma viral load or CD4⁺ T-cell loss compared with the remaining controls.

During the acute phase of the infection, most monkeys immunized with either the DNA or MVA vectors in study A or with the DNA/CRL1005–MVA vector prime–boost combination in study B exhibited an acute CD4⁺ T-cell lymphopenia similar to that in the control animals. The peak loads of plasma virus in these immunized animals were also indistinguishable from those in the controls. However, by about 70 days after challenge, many of the immunized monkeys exhibited some evidence of a positive immunization benefit, as manifested by control of viraemia and recovery of CD4⁺ T-cell counts. This was particularly noticeable in the monkeys that had been immunized with four doses of the DNA/CRL1005 vaccine.

Nonetheless, the greatest degree of attenuation of the SHIV infection was provided in both studies by immunization with the Ad5-vector vaccine, either alone or as a boost after priming with the DNA vector formulated in CRL1005 (Fig. 2). With a single exception, the CD4⁺ T-cell counts in these vaccinated monkeys remained above 500 cells μL^{-1} at all times and recovered to around 1,000 cells μL^{-1} by 70 days after challenge. In contrast to the other vaccine groups, the peak viraemia levels in the Ad5-vaccinated monkeys were 10- to 50-fold lower than those in the unimmunized monkeys, and all controlled the chronic loads of plasma virus to levels close to the assay detection limit (500 viral RNA copies per ml). One Ad5-immunized animal in study B, 98C040, did exhibit a somewhat prolonged peak viraemia with an associated acute loss of CD4⁺ T cells. Yet, by 120 days after challenge, this monkey also manifested both a noted control of viraemia and recovery of CD4⁺ T cells.

The observations of viraemia and CD4⁺ T cells that were made with peripheral blood samples were confirmed by *in situ* staining analyses of lymph node biopsies (data not shown).

Serum samples collected after the challenge were assayed for neutralizing antibody against SHIV 89.6P. For most of the immunized animals, and for all of the Ad5-immunized monkeys, measurable neutralizing antibody responses were not detected until 42 days after challenge (data not shown). This is much later than the relatively rapid post-challenge neutralizing antibody responses reported in animals from previous studies that had been immunized with vectors expressing the viral *env* gene^{9,10}. Also, the elicitation of neutralizing antibodies after challenge in our studies was not dependent on the type of vaccine used. These observations support the conclusion that the early mitigation of viral infection reported here was due solely to the vaccine-elicited gag-specific cellular immune responses. Any potentially confounding effects of the initial post-challenge neutralizing antibody response were controlled by avoiding the use of the *env* gene for immunization.

Immunodeficiency-related symptoms, such as prolonged weight loss, opportunistic infections and chronic diarrhoea, have been observed in five of the six unimmunized control monkeys from study A to date (~1 year after challenge). These five monkeys have been euthanized owing to their deteriorating physical condition (according to the recommendation of the attending veterinarians and under institutional guidelines). In contrast, the immunized animals, with two exceptions, generally have remained healthy during this time. The two monkeys that did not remain healthy (T282 and 088R from the DNA and the DNA/MPL/Alum vaccine

groups, respectively) exhibited immunodeficiency-related symptoms beginning about six months after challenge, and we euthanized T282 at day 290. These two animals had developed relatively low immune responses after immunization and had poorly controlled viraemia after challenge. The monkeys immunized with the Ad5 vector have remained completely healthy throughout the post-challenge interval.

The unimmunized control monkeys from study B have all shown various degrees of immunodeficiency-related disease during the six months after challenge. Four of the eight controls have been euthanized owing to their deteriorating health. Two of the euthanized controls were *Mamu-A*01* positive while two were not, again suggesting that the expression of this specific MHC class I allele does not influence the course of infection or disease after challenge with SHIV 89.6P. None of the vaccinated monkeys from study B have yet exhibited any signs of immunodeficiency or suffered any consistent weight loss.

The results of the two studies reported here demonstrate that vaccine-elicited cellular immune responses against SIV gag can successfully attenuate infection and mitigate disease progression in a stringent model of monkey challenge using a highly pathogenic SHIV strain. The most notable vaccine-mediated effects were seen in those animals that had been immunized with the replication-incompetent Ad5 vector, either alone or as a booster inoculation after priming with a DNA vector formulated in CRL1005. The best pre-challenge immune responses in both studies were exhibited by the Ad5-immunized monkeys. After virus challenge, these Ad5-immunization-mediated responses were effective in reducing the level of peak viraemia, minimizing loss of CD4⁺ T cells during the infection's acute phase and maintaining low levels of viraemia during the chronic phase, thereby permitting almost complete recovery of CD4⁺ T cells. It is of interest that these beneficial post-challenge effects were seen in both immunization groups that involved the use of the Ad5 vector, although the vaccine-elicited cellular immune response in the heterologous prime–boost cohort (study B) was almost fivefold greater than that seen in the cohort immunized with the Ad5 vector alone (study A). Boosting with the MVA vector after DNA/CRL1005 priming was not nearly as effective in mitigating the effects of virus challenge; yet, the pre-challenge immune responses seemed similar to those elicited by immunization with only the Ad5 vector. Care must be used in making such quantitative comparisons across two independent studies. Nonetheless, these observations suggest that the Ad5 vector was superior with regard to both the quantity and, possibly, the quality of the elicited immune response. Further immunization and challenge studies are in progress to evaluate the nature of the Ad5-vector-mediated systemic and local (lymph node) virus-specific cellular immune response immediately before and after challenge in an attempt to establish an unequivocal immunological correlate of the antiviral effect. The relevance of the SHIV 89.6P monkey challenge model system used in these studies to human HIV-1 infection is not firmly established. The potential of the vaccines studied in this model cannot presently be extrapolated to predict human immunogenicity or efficacy. This assessment awaits the outcome of ongoing safety and immunogenicity clinical trials. □

Methods

Animals

Rhesus monkeys (*Macaca mulatta*) were maintained in accordance with the institutional animal care protocols of Merck Research Laboratories, New Iberia Research Center, and Primedica. Monkeys were typed for expression of the *Mamu-A*01* allele according to PCR sequence-specific primers methods as previously described¹⁷. Animals that were positive for the PCR reaction were subsequently confirmed by DNA sequencing of the amplified region. Twelve weeks after the final immunizations, monkeys were challenged intravenously with 50 MID₅₀ of cell free SHIV-89.6P¹⁶. Monkeys were monitored for clinical signs of disease progression and cared for under the guidelines established by the National Institutes of Health (NIH) 'Guide for the Care and Use of Laboratory Animals'.

Vaccines

The codon-optimized SIVmac239 *gag* gene was constructed by annealing a series of overlapping oligonucleotides. The authenticity of the synthetic gene was confirmed by DNA sequencing. The gene was cloned into V1R vector, which has been described previously^{18,19}. Plasmid DNA vaccines were formulated either in PBS alone, or in PBS containing the selected adjuvants: 7.5 mg of a nonionic block copolymer (CRL1005, CytRx); or 0.5 mg of monophosphoryl lipid A (MPL, Corixa) absorbed onto 0.7 mg of aluminium, as aluminium hydroxyphosphate (Adju-Phos, HCl Biosector). For study B, a modification of the DNA/CRL1005 was employed on the basis of ongoing studies that suggested this formulation provided improved immune responses compared with the initial formulation. For this formulation, 5 mg of DNA was added to 7.5 mg of CRL1005 mixed with 0.75 mM benzalkonium chloride (Ruger). Before immunization, the formulation was warmed slowly to room temperature from a frozen stock.

The synthetic SIV *gag* gene was also used for construction of the MVA and adenoviral vectors. The method for generation of recombinant MVA has been described elsewhere²⁰. Briefly, the SIV *gag* gene was cloned into the pSC59 shuttle vector. This plasmid was designed to insert the transgene fragment into a viral thymidine kinase region, and to drive the transgene from a synthetic early/late promoter. Recombinant virus was identified and subsequently cloned again from single plaques six times with confirmation of maintenance of transgene by detecting expression by immunostaining *gag*.

The adenoviral vector was based on a serotype 5 adenovirus that had been rendered incompetent to replicate by deletion of the E1 and E3 viral genes, and was propagated subsequently in E1-expressing 293 cells. Recombinant adenovirus expressing the codon-optimized SIV *gag* gene was prepared as follows. The adenoviral shuttle vector, pHCMVIBGHpA1, contains Ad5 sequences of nucleotides 1–341 and 3,534–5,798 and an expression cassette containing the human cytomegalovirus (HCMV) promoter with intron A and the bovine growth hormone polyadenylation signal. The codon-optimized full-length SIV *gag* was digested from a parental plasmid with *Bgl*II and then cloned into a *Bgl*II-digested shuttle vector, pHCMVIBGHpA1. The shuttle vector was recombined with an adenoviral backbone plasmid, pAd-ΔE1E3, which contains the complete Ad5 genome except the E1 and E3 regions, to generate a pre-adenoviral plasmid, pAd5-SIV*gag*. This plasmid was linearized by digestion with *Pac*I and transfected into 293 cells to generate Ad5-SIV*gag*. The recombinant adenovirus was grown in large quantities by multiple rounds of amplification in 293 cells. The virus was purified by caesium chloride gradient centrifugation. Viral DNA was extracted by proteinase K digestion and confirmed for genomic integrity by restriction enzyme digestion of the viral genome, and PCR and DNA sequencing of the transgene cassette. SIV *gag* expression was verified by western blot analysis of COS and 293 cells infected with the virus. Virus quantities were expressed in terms of viral particles, based on determination of the nucleic acid content from purified virus.

Tetramer staining

The methods for preparing p11CM tetramer reagent and staining lymphocytes within whole blood have been described elsewhere². Samples were analysed by flow cytometry on a FACScalibur (Becton Dickinson). Gated CD3⁺ CD8⁺ T cells were examined for staining with tetrameric Mamu-A*01–p11CM complex. The CTL epitope p11CM has been reported as a dominant epitope in SIV *gag* restricted to the Mamu-A*01 MHC class I allele. For each sample, 30,000 gated CD3⁺ CD8⁺ T-cell events were collected and analysed on CellQuest program (Becton Dickinson), and the results presented as per cent tetramer-positive cells within CD3⁺-and-CD8⁺-positive events.

Received 30 August; accepted 18 October 2001.

- Koup, R. A. *et al.* Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome. *J. Virol.* **68**, 4650–4655 (1994).
- Borrow, P., Lewicki, H., Hahn, B. H., Shaw, G. M. & Oldstone, M. B. Virus-specific CD8⁺ cytotoxic T-lymphocyte activity associated with control of viremia in primary human immunodeficiency virus type 1 infection. *J. Virol.* **68**, 6103–6110 (1994).
- Kuroda, M. J. *et al.* Emergence of CTL coincides with clearance of virus during primary simian immunodeficiency virus infection in rhesus monkeys. *J. Immunol.* **162**, 5127–5133 (1999).
- Klein, M. R. *et al.* Associations between HLA frequencies and pathogenic features of human immunodeficiency virus type 1 infection in seroconverters from the Amsterdam cohort of homosexual men. *J. Infect. Dis.* **169**, 1244–1249 (1994).
- Musey, L. *et al.* Cytotoxic-T-cell responses, viral load, and disease progression in early human immunodeficiency virus type 1 infection. *N. Engl. J. Med.* **337**, 1267–1274 (1997).
- Pontesilli, O. *et al.* Longitudinal analysis of human immunodeficiency virus type 1-specific cytotoxic T lymphocyte responses: a predominant *gag*-specific response is associated with nonprogressive infection. *J. Infect. Dis.* **178**, 1008–1018 (1998).
- Ogg, G. S. *et al.* Longitudinal phenotypic analysis of human immunodeficiency virus type 1-specific cytotoxic T lymphocytes: correlation with disease progression. *J. Virol.* **73**, 9153–9160 (1999).
- Gallimore, A. *et al.* Early suppression of SIV replication by CD8⁺ nef-specific cytotoxic T cells in vaccinated macaques. *Nature Med.* **1**, 1167–1173 (1995).
- Barouch, D. H. *et al.* Control of viremia and prevention of clinical AIDS in rhesus monkeys by cytokine-augmented DNA vaccination. *Science* **290**, 486–492 (2000).
- Amara, R. R. *et al.* Control of a mucosal challenge and prevention of AIDS by a multiprotein DNA/MVA vaccine. *Science* **292**, 69–74 (2001).
- Schmitz, J. E. *et al.* Control of viremia in simian immunodeficiency virus infection by CD8⁺ lymphocytes. *Science* **283**, 857–860 (1999).
- Newman, M. J., Actor, J. K., Balusubramanian, M. & Jagannath, C. Use of nonionic block copolymers in vaccines and therapeutics. *Crit. Rev. Ther. Drug Carrier Syst.* **15**, 89–142 (1998).
- Miller, M. D., Yamamoto, H., Hughes, A. L., Watkins, D. I. & Letvin, N. L. Definition of an epitope and MHC class I molecule recognized by *gag*-specific cytotoxic T lymphocytes in SIVmac-infected rhesus monkeys. *J. Immunol.* **147**, 320–329 (1991).

- Allen, T. M. *et al.* Characterization of the peptide binding motif of a rhesus MHC class I molecule (Mamu-A*01) that binds an immunodominant CTL epitope from simian immunodeficiency virus. *J. Immunol.* **160**, 6062–6071 (1998).
- Egan, M. A. *et al.* Use of major histocompatibility complex class I/peptide/β2 M tetramers to quantitate CD8⁺ cytotoxic T lymphocytes specific for dominant and nondominant viral epitopes in simian-human immunodeficiency virus-infected rhesus monkeys. *J. Virol.* **73**, 5466–5472 (1999).
- Reimann, K. A. *et al.* An *env* gene derived from a primary human immunodeficiency virus type 1 isolate confers high in vivo replicative capacity to a chimeric simian/human immunodeficiency virus in rhesus monkeys. *J. Virol.* **70**, 3198–3206 (1996).
- Knapp, L. A., Lehmann, E., Piekarczyk, M. S., Urvater, J. A. & Watkins, D. I. A high frequency of Mamu-A*01 in the rhesus macaque detected by polymerase chain reaction with sequence specific primers and direct sequencing. *Tissue Antigens* **50**, 657–661 (1997).
- Montgomery, D. L. *et al.* Heterologous and homologous protection against influenza A by DNA vaccination: optimization of DNA vectors. *DNA Cell Biol.* **12**, 777–783 (1993).
- Fu, T. M., Friedman, A., Ulmer, J. B., Liu, M. A. & Donnelly, J. J. Protective cellular immunity: cytotoxic T-lymphocyte responses against dominant and recessive epitopes of influenza virus nucleoprotein induced by DNA immunization. *J. Virol.* **71**, 2715–2721 (1997).
- Wyatt, L. S., Shors, S. T., Murphy, B. R. & Moss, B. Development of a replication-deficient recombinant vaccinia virus vaccine effective against parainfluenza virus 3 infection in an animal model. *Vaccine* **14**, 1451–1458 (1996).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We gratefully acknowledge K. Wright of Laboratory Animal Resources, Merck Research Laboratories, R. Druilhet and R. Colligan of New Iberia Research Center, and M. Hevy and K. Manson of Primatec for their contributions to this research. We are also thankful for assistance from B. Moss and V. Hirsch of the NIH in constructing the recombinant MVA virus.

Correspondence and requests for materials should be addressed to E.A.E. (e-mail: emilio_emini@merck.com).

Eventual AIDS vaccine failure in a rhesus monkey by viral escape from cytotoxic T lymphocytes

Dan H. Barouch^{*}, Jennifer Kunstman[†], Marcelo J. Kuroda^{*}, Jörn E. Schmitz^{*}, Sampa Santra^{*}, Fred W. Peyerl^{*}, Georgina R. Krivulka^{*}, Kristin Beaudry^{*}, Michelle A. Lifton^{*}, Darci A. Gorgone^{*}, David C. Montefiori[‡], Mark G. Lewis[§], Steven M. Wolinsky[†] & Norman L. Letvin^{*}

^{*} Department of Medicine, Harvard Medical School, Beth Israel Deaconess Medical Center, Research East Room 113, 330 Brookline Avenue, Boston, Massachusetts 02215, USA

[†] Department of Medicine, Northwestern University Medical School, Chicago, Illinois 60611, USA

[‡] Department of Surgery, Duke University Medical Center, Durham, North Carolina 27710, USA

[§] Southern Research Institute, Frederick, Maryland 21701, USA

Potent virus-specific cytotoxic T lymphocyte (CTL) responses elicited by candidate AIDS vaccines have recently been shown to control viral replication and prevent clinical disease progression after pathogenic viral challenges in rhesus monkeys^{1–4}. Here we show that viral escape from CTL recognition can result in the eventual failure of this partial immune protection. Viral mutations that escape from CTL recognition have been previously described in humans infected with human immunodeficiency virus (HIV)^{5–10} and monkeys infected with simian immunodeficiency virus (SIV)^{11–13}. In a cohort of rhesus monkeys that were vaccinated and subsequently infected with a pathogenic hybrid simian–human immunodeficiency virus (SHIV), the frequency of viral sequence mutations within CTL epitopes correlated with the level of viral replication. A single nucleotide mutation within an immunodominant *Gag* CTL epitope in an animal with undetect-

able plasma viral RNA resulted in viral escape from CTLs, a burst of viral replication, clinical disease progression, and death from AIDS-related complications. These data indicate that viral escape from CTL recognition may be a major limitation of the CTL-based AIDS vaccines that are likely to be administered to large human populations over the next several years.

We recently described the immunogenicity and protective efficacy in rhesus monkeys (*Macaca mulatta*) of a vaccine strategy that used *env/gag* plasmid DNA vaccines augmented by an interleukin-2 (IL-2)–immunoglobulin cytokine fusion protein or an IL-2–immunoglobulin plasmid¹. After challenge with the pathogenic SHIV-89.6P (ref. 14), control animals developed high plasma viral RNA levels and exhibited a profound depletion of CD4⁺ T lymphocytes, clinical disease progression, and mortality. In contrast, eight of eight animals that received the cytokine-augmented DNA vaccines developed potent virus-specific CTL responses and maintained effective control of viral replication with no evidence of clinical disease progression for 140 days after challenge¹. Seven of these animals have also maintained durable control of viral replication, preserved CD4⁺ T lymphocytes, and a stable clinical status for more than 2 years after challenge (D.H.B. and N.L.L., unpublished data). One monkey that received the DNA vaccines plus IL-2–immunoglobulin protein (monkey 798), however, had an eventual vaccine failure. After the resolution of primary viraemia (weeks 0–10), this animal initially demonstrated effective control of viral replication with undetectable levels of plasma viral RNA (<500 copies per ml; weeks 10–20) (Fig. 1a). This monkey then exhibited a breakthrough of viral replication (week 24), loss of CD4⁺ T lymphocytes (week 36), symptomatic clinical disease (week 44), and death from an AIDS-like syndrome (week 52) (Fig. 1a, b). Necropsy of this animal revealed severe lymphopaenia in lymph nodes, thymus and spleen, and gliosis in the central nervous system.

Because monkey 798 expressed the major histocompatibility complex (MHC) class I molecule Mamu-A*01, we prospectively monitored the CD8⁺ T lymphocyte responses specific for several Mamu-A*01-restricted CTL epitopes in this animal. Longitudinal analysis of immune responses from this monkey after resolution of primary viraemia revealed a selective decline of the CD8⁺ T lymphocyte response specific for the immunodominant Mamu-A*01-restricted SIV Gag p11C epitope (CTPYDINQM)^{15,16}, as determined by p11C tetramer staining of freshly isolated CD8⁺ T lymphocytes^{17,18}. At week 28, after the breakthrough of viral replication and before the loss of CD4⁺ T lymphocytes, this animal exhibited a dramatic decline of its p11C tetramer response (Fig. 1c). Interestingly, tetramer responses to the subdominant Mamu-A*01-restricted HIV-1 Env p41A (YAPPTGQI) and SIV Pol p68A (STPPLVRLV) epitopes¹⁹ increased markedly at this time (Fig. 1d, e). The resurgent virus population in monkey 798 thus seemed to stimulate the selective expansion of the Env p41A-specific and Pol p68A-specific but not the Gag p11C-specific CTL populations. Neutralizing antibody responses to SHIV-89.6P, as measured by MT-2 cell killing assays²⁰, also increased slightly during this period, probably stimulated by the increased viral replication (Fig. 1f).

We next investigated the frequency of viral sequence mutations within CTL epitopes in the cohort of control and vaccinated monkeys after SHIV-89.6P challenge¹. At the setpoint of infection at week 14, we sequenced 500-base-pair (bp) regions of *gag* and *env* containing four Mamu-A*01-restricted CTL epitopes from 8–12 viral clones per animal²¹. Viral sequence analysis revealed high frequencies of background sequence diversity within the *gag* and *env* genes that correlated with the mean setpoint viral RNA levels of the animals ($P = 0.003$, using a two-sided Spearman rank correlation test) (Fig. 2a). Similar high frequencies of mutations were observed within CTL epitopes and correlated with mean setpoint viral RNA levels ($P = 0.001$) (Fig. 2b). Interestingly, mutations within Mamu-A*01-restricted CTL epitopes were also found in the Mamu-A*01-negative animals (Fig. 2, open symbols). These

observations suggest that the level of viral replication, rather than immune selection pressure, drove the random emergence of viral sequence mutations within CTL epitopes.

Longitudinal analysis of *gag*, *env* and *pol* sequences from monkey 798 revealed the rapid emergence of a single nucleotide mutation within the immunodominant Gag p11C CTL epitope (Fig. 3).

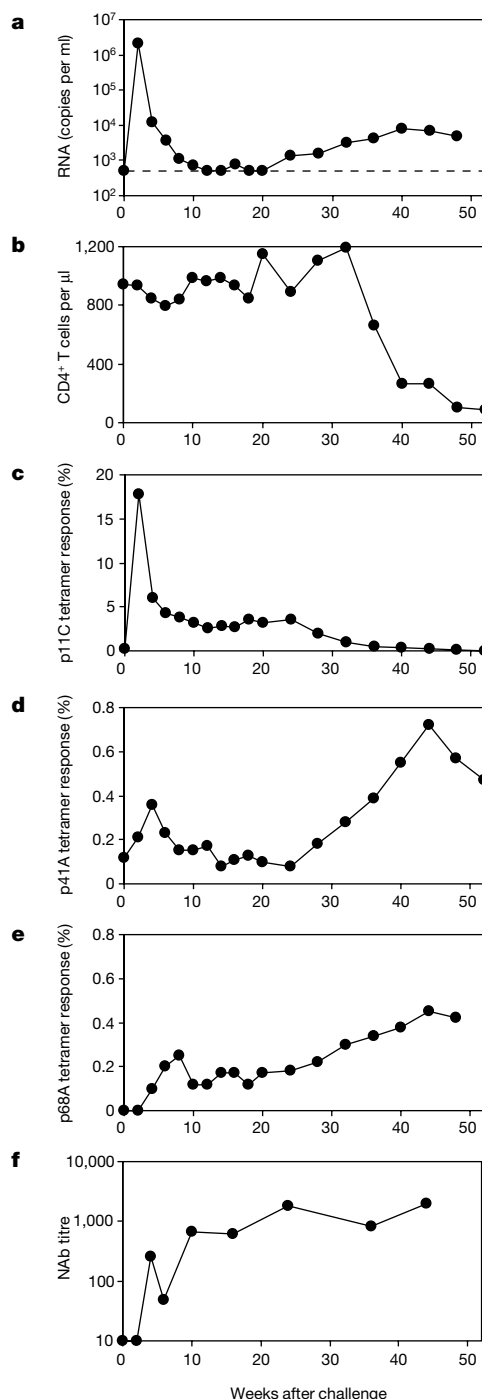


Figure 1 Clinical and immunologic parameters of monkey 798. This Mamu-A*01-positive monkey was vaccinated with *gag/env* DNA vaccines plus IL-2–immunoglobulin protein and challenged with SHIV-89.6P. Longitudinal analyses of (a) levels of plasma viral RNA, (b) CD4⁺ T lymphocyte counts, (c) Gag p11C tetramer responses, (d) Env p41A tetramer responses, (e) Pol p68A tetramer responses, and (f) SHIV-89.6P-specific neutralizing antibody (NAb) responses are shown. The dashed line in a represents the detection limit for plasma viral RNA of 500 copies per ml. The tetramer responses represent the percentage of gated CD3⁺ CD8⁺ T lymphocytes that bound tetramers.

Between weeks 14 and 20, just before the breakthrough of viral replication, there was a cytidine to thymidine mutation that resulted in a threonine to isoleucine substitution at position 2 of the p11C epitope (CTPYDINQM to CIPYDINQM). This mutation was found in no (0 of 8) viral clones from plasma samples at week 14, but was found in all (53 of 53) viral clones isolated from plasma samples at weeks 20, 24, 28, 36 and 44. This mutation was found in no (0 of 15) viral clones isolated from the original SHIV-89.6P challenge stock. The rapid and sustained replacement of the entire viral population by a 'quasispecies' containing this mutation demonstrates the high level of immune selection pressure exerted by p11C-specific CTLs. Although the level of viral replication drove the initial emergence of CTL epitope mutations, immune selection pressure seemed to determine the extent to which the mutant virus replaced the viral population.

As compared with the wild-type p11C peptide, the mutant p11C peptide had more than 100-fold lower affinity for Mamu-A*01 as determined by cellular peptide binding assays (Fig. 4a). Moreover, recombinant Mamu-A*01 complexes refolded *in vitro* around the mutant p11C peptide were found to be unstable and rapidly dissociated (data not shown). In addition, the mutant p11C peptide had more than 1,000-fold lower recognition by CD8⁺ T lymphocytes from Mamu-A*01-positive SHIV-89.6P-infected rhesus monkeys, as determined by interferon- γ ELISPOT assays (Fig. 4b). CD8⁺ T lymphocytes from monkey 798 recognized neither the wild-type nor the mutant p11C peptide before its death. These data demonstrate that the mutant p11C peptide escaped efficient recognition by p11C-specific CTLs.

The frequencies of viral mutations within five subdominant

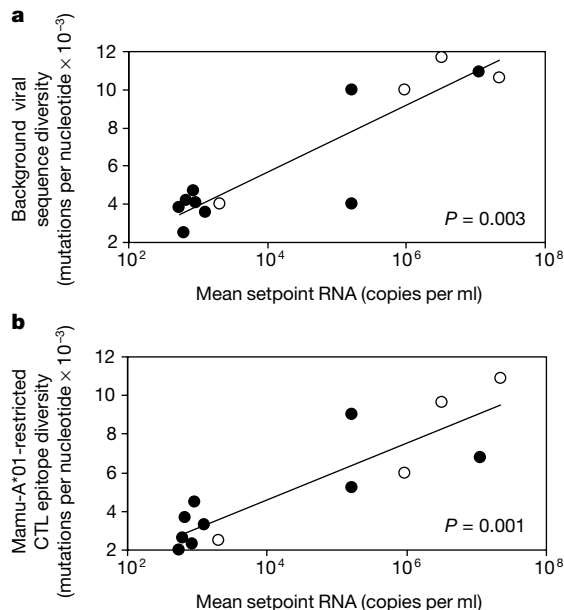


Figure 2 Background and CTL epitope sequence mutations. In a cohort of control and vaccinated rhesus monkeys infected with SHIV-89.6P, plasma viral sequence analyses were performed on 8–12 viral clones per animal for *gag* and *env* at week 14 after infection. **a**, The frequencies of background viral sequence mutations within 500-bp regions of *gag* and *env* are shown as a function of the mean setpoint viral RNA levels of the animals ($P = 0.003$). **b**, The frequencies of viral sequence mutations within the Mamu-A*01-restricted Gag p11C, Env p41A, Gag LW9 and Gag QI9 CTL epitopes are also shown as a function of the mean setpoint viral RNA levels ($P = 0.001$). The mean frequency of coding mutations per nucleotide is reported for each animal. Filled symbols represent the Mamu-A*01-positive animals; open symbols represent the Mamu-A*01-negative animals. Statistical analyses were performed by two-sided Spearman rank correlation tests with a 95% confidence interval.

Mamu-A*01-restricted CTL epitopes^{19,22}, Gag LW9 (LSPRTLNAW), Gag QI9 (QNPIPVGNI), Env p41A (YAPPITGQI), Pol QV9 (QVPKFHLPV) and Pol p68A (STPPLVRLV), were low in this animal (Fig. 3). In fact, the mutation frequencies within subdominant CTL epitopes were in general comparable to the frequencies of background sequence mutations within *gag*, *env* and *pol* (data not shown), suggesting that these mutations were under weak immune selection pressure. The notable exception to this was the late emergence at week 44 of a Pol p68A mutation that resulted in a threonine to isoleucine substitution at position 2 of this epitope (STPPLVRLV to SIPPLVRLV). This mutant virus replaced most of the viral population in the animal. The mutant p68A peptide also had more than 100-fold lower affinity for Mamu-A*01 and more than 1,000-fold lower recognition by CD8⁺ T lymphocytes as determined by peptide binding and ELISPOT assays, indicating that it escaped efficient recognition by p68A-specific CTLs (data not shown). Interestingly, monkey 798 exhibited a marked rise in p68A tetramer staining after the p11C mutation (Fig. 1e). The late p68A mutation therefore occurred in the setting of a prolonged

Gag p11C (181–189)

	C T P Y D I N Q M
week 0	- - - - - (15/15)
week 14	- - - - - (8/8)
week 20	- I - - - - - (10/10)
week 24	- I - - - - - (11/11)
week 28	- I - - - - - (11/11)
week 36	- I - - - - - (11/11)
week 44	- I - - - - - (10/10)

Gag LW9 (149–157)

	L S P R T L N A W
week 0	- - - - - (15/15)
week 14	- - - - - (8/8)
week 20	- - - - - (10/10)
week 24	- - - - - (11/11)
week 28	- - - - - (11/11)
week 36	- - - - - (11/11)
week 44	- - - - - (10/10)

Gag QI9 (254–262)

	Q N P I P V G N I
week 0	- - - - - (15/15)
week 14	- - - - - (8/8)
week 20	- - - - - (10/10)
week 24	- - - - - (11/11)
week 28	- - - - - (11/11)
week 36	- - - - - (10/11)
week 44	- - - - - (10/10)

Env p41A (431–439)

	Y A P P I T G Q I
week 0	- - - - - (19/22)
week 14	- - - - - (1/22)
week 20	- - - - - (1/22)
week 24	- - - - - (1/22)
week 28	- - - - - (1/11)
week 36	- - - - - (1/11)
week 44	- - - - - (1/11)

Pol QV9 (588–596)

	Q V P K F H L P V
week 0	- - - - - (14/14)
week 14	- - - - - (8/8)
week 20	- - - - - (11/11)
week 24	- - - - - (12/12)
week 28	- - - - - (10/11)
week 36	- - - - - (1/11)
week 44	- - - - - (8/8)

Pol p68A (621–629)

	S T P P L V R L V
week 0	- - - - - (14/14)
week 14	- - - - - (7/8)
week 20	- - - - - (1/8)
week 24	- - - - - (11/11)
week 28	- - - - - (11/12)
week 36	- - - - - (8/11)
week 44	- - - - - (1/12)

Figure 3 Emergence of CTL epitope mutations in virus isolated from monkey 798.

Longitudinal analyses of plasma viral sequences were performed on 8–12 viral clones per time point for *gag*, *env* and *pol*. The deduced amino acid sequences of the Mamu-A*01-restricted immunodominant Gag p11C epitope and the subdominant Gag LW9, Gag QI9, Env p41A, Pol QV9 and Pol p68A epitopes are shown. The number of mutations and the number of viral clones analysed per time point are shown in parentheses.

resurgence of viral replication and increased immune selection pressure focused on this epitope.

Thus, in an animal with effective vaccine-elicited immune control of viral replication, a single nucleotide mutation in an immunodominant CTL epitope was associated with a breakthrough of viral replication and clinical disease progression. This dramatic viral escape from immune control occurred in the setting of undetectable levels of plasma viral RNA, preserved functional CD4⁺ T lymphocytes, CD8⁺ T lymphocyte responses to a diversity of viral epitopes, and high neutralizing antibody titres. It is, however, important to note that the levels of plasma viral RNA associated with the late resurgence of viral replication and progressive loss of CD4⁺ T lymphocytes (weeks 24–48) were considerably lower than the viral RNA levels associated with primary viraemia in this monkey (weeks 0–10; Fig. 1a, b). Thus, although our findings clearly implicate this single viral nucleotide substitution in the clinical decompensation of this monkey, we cannot exclude the possibility that other mutations may also have contributed to the pathogenicity of this virus.

Viral mutations within CTL epitopes in *nef* and *tat* rather than *gag* have been previously described in SIV-infected rhesus monkeys^{11,12}. The clinical decline associated with the mutation of the immunodominant Gag epitope here reflects the functional significance of this epitope-specific immune response and the immune selection pressure that is focused on this epitope. These data extend prior studies demonstrating the critical importance of CTLs in maintaining control of viral replication^{23,24} and the emergence of eventual breakthroughs in viral replication after challenge of vaccinated monkeys²⁵. The decline of p11C-specific tetramer responses after the p11C mutation demonstrates that a constant antigenic stimulus is required for maintaining high levels of epitope-specific CTLs. This finding is consistent with prior studies of CTL responses in patients after highly active antiretroviral treatment^{26,27}.

Although differences certainly exist between the pathogenic consequences of SHIV-89.6P infection of rhesus monkeys and HIV-1 infection of humans, our findings suggest that viral escape from CTL recognition will prove to be a general mechanism for loss of immune control of viral replication in infected individuals. In our study, mutations within CTL epitopes occurred frequently in a large cohort of monkeys. It is clear that even animals with undetectable levels of plasma viral RNA have sufficient ongoing viral replication to generate CTL escape mutations. Moreover, after CTL escape occurs at an immunodominant epitope, the resurgent viral replication and increased immune selection pressure on subdominant epitopes can lead to additional immune escape at these epitopes and a subsequent rapid clinical decline.

Current HIV-1 vaccine strategies that are based on plasmid DNA and recombinant live vectors have achieved only partial protection rather than sterilizing immunity against viral challenges in rhesus monkeys^{1–4}. Our study has several clear implications for these CTL-based vaccine strategies. It will be important to develop vaccines that elicit the greatest breadth of CTL responses and that drive viral replication to the lowest possible levels after infection. It will also be critical to develop immunogens that elicit broadly reactive neutralizing antibody responses. It is sobering to find that a single point mutation within the virus can initiate a cascade of events resulting in clinical vaccine failure and death. In fact, until a vaccine is developed that induces sterilizing immunity, CTL escape may be a general mechanism of vaccine failure and a major practical limitation of the current generation of CTL-based AIDS vaccines. □

Methods

CD4⁺ T lymphocyte counts and viral RNA levels

Counts of CD4⁺ T lymphocytes were determined by monoclonal antibody staining and flow cytometry. Plasma viral RNA levels were measured by an ultrasensitive branched DNA amplification assay with a detection limit of 500 copies per ml (Bayer Diagnostics).

Tetramer staining

PE-labelled tetrameric Mamu-A*01–peptide complexes (0.2 µg) were used in conjunction with fluorescein isothiocyanate (FITC)-labelled anti-human CD8α (Leu2a; Becton Dickinson), ECD-labelled anti-human CD8αβ (2S8-5H7; Beckman Coulter), and APC-labelled anti-rhesus-monkey CD3 (FN18) monoclonal antibodies to stain epitope-specific CD8⁺ T cells as described^{17,18}. Whole blood from the monkeys (100 µl) was stained with these reagents, lysed, washed and fixed. Samples were analysed by four-colour flow cytometry on a Becton Dickinson FACScalibur, and gated CD3⁺ CD8⁺ T cells were examined for staining with tetrameric Mamu-A*01–peptide complexes.

Neutralizing antibody assays

Reduction of virus-induced cytopathic killing of MT-2 cells by SHIV-89.6P-specific antibodies was assessed as described²⁰. Fifty microlitres of cell-free virus containing 500 TCID₅₀ (50% tissue culture infective dose) grown in human peripheral blood mononuclear cells (PBMCs) was added to multiple dilutions of test plasma in 100 µl of growth media in triplicate. These mixtures were incubated for 1 h before the addition of 5 × 10⁴ MT-2 cells. Infection led to extensive syncytium formation and viral-induced cell killing in 4–6 days in the absence of neutralizing antibodies. Neutralization titres were calculated as the reciprocal dilution of plasma required to protect 50% of cells from virus-induced killing.

Viral sequencing

Viral sequence analyses of 500-bp regions of *gag*, *env* and *pol* were performed as described²¹. Virus from frozen plasma samples was isolated by centrifugation at 25,000g for 1 h and lysed with 48% guanidine thiocyanate, 1.4% dithiothreitol, 1% N-lauroylsarcosine and 1% sodium citrate. RNA was precipitated with isopropanol and solubilized. First-strand complementary DNA was synthesized with reverse transcriptase using the SIV *gag* primer TGTGTTCTGCTCTTAAGCTTTTGTAG or the SIV *pol* primer ATGCCATGAGAAATGCTTCCA. First-strand cDNA was synthesized during PCR amplification for HIV-1 *env*. Initial PCR amplification used the following primers: *gag*-fwd, ACCTAGTGGTGGAAACAGGAACAG; *gag*-rev, TGTGTTCTGCTCTTAAGCTTTTGTAG; *env*-fwd, CCAATCCCATACATTATTGT; *env*-rev, ATAGTGCTTCTGCTGCTCCCAAGAACC; *pol*-fwd, TTCTCAGTCAGGAACAAGAAGGATG; *pol*-rev, ATGCCATGAGAAATGCTTCCA. Secondary nested PCR amplification used the following primers: *gag*-fwd, AGCACCATCTAGTGGCAGAGGA; *gag*-rev, GAAATGGCTCTTTGGCCCTT; *env*-fwd, CUACUACUACUATCCATCCAGTCTAGCAGAGAAGA; *env*-rev, CAUCAUCAUCAUATGAGGGACAATTCTAGAAGT; *pol*-fwd, CAAGAAGGCAAGCCATTAGAAGC; *pol*-rev, TCTGTGATATATCTGCTTCCCTTC. The amplified fragments were cloned

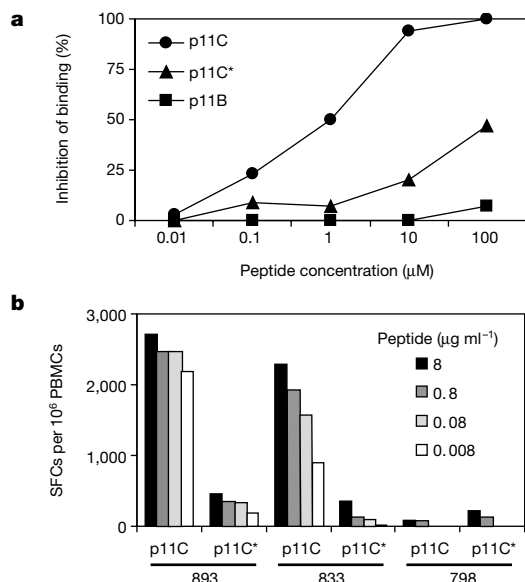


Figure 4 The mutant p11C peptide escapes recognition by p11C-specific CTL. **a**, Cellular peptide-binding assays of the wild-type p11C peptide and the mutant p11C peptide (p11C*) for the MHC class I molecule Mamu-A*01. The p11B peptide (ALSEGCTPYDIN) is an irrelevant peptide control. **b**, Functional IFN-γ ELISPOT assays using PBMCs from monkeys 893, 833 and 798 at 52 weeks after challenge. All of these monkeys expressed the Mamu-A*01 MHC class I allele. PBMCs from these animals were stimulated with either the wild-type p11C peptide or the mutant p11C peptide (p11C*) at 8, 0.8, 0.08 and 0.008 µg ml⁻¹. Mean spot-forming cells (SFCs) per 10⁶ PBMCs are shown. The mean responses of assays performed in triplicate are shown.

into pCRII-TOPO (Invitrogen) or pAMPI (Stratagene), and individual transformed colonies were subjected to T7/SP6 dideoxy sequencing.

Peptide-binding assays

Peptide-binding assays were performed by measuring inhibition of iodinated p11C peptide binding to 721.221 cells that express the MHC class I molecule Mamu-A*01. 2×10^6 cells were incubated overnight at 26 °C with 3 $\mu\text{g ml}^{-1}$ human $\beta_2\text{m}$; 1.5×10^5 c.p.m. iodinated p11C peptide and serial log dilutions of unlabelled test peptides were then incubated with washed cells for 4 h at 20 °C. Cells were then washed three times, and the radioactivity of the cell pellet was measured with a scintillation counter. Per cent inhibition of binding by the test peptides was calculated as $1 - (\text{c.p.m. with competitor peptide}/\text{c.p.m. without competitor peptide})$.

ELISPOT assays

We coated 96-well multiscreen plates overnight with 100 μl per well of 10 $\mu\text{g ml}^{-1}$ anti-human interferon- γ (IFN- γ) (B27; BD Pharmingen) in endotoxin-free Dulbecco's PBS (D-PBS). The plates were then washed three times with D-PBS containing 0.25% Tween-20 (D-PBS/Tween), blocked for 2 h with D-PBS containing 5% FBS at 37 °C, washed three times with D-PBS/Tween, rinsed with RPMI 1640 containing 10% FBS to remove the Tween-20, and incubated with 8, 0.8, 0.08 or 0.008 $\mu\text{g ml}^{-1}$ peptide and 2×10^5 PBMCs in triplicate in 100- μl reaction volumes. Following an 18-h incubation at 37 °C, the plates were washed nine times with D-PBS/Tween and once with distilled water. The plates were then incubated with 2 $\mu\text{g ml}^{-1}$ biotinylated rabbit anti-human IFN- γ (Biosource) for 2 h at room temperature, washed six times with Coulter Wash (Beckman-Coulter), and incubated for 2.5 h with a 1:500 dilution of streptavidin-AP (Southern Biotechnology). After five washes with Coulter Wash and one with PBS, the plates were developed with NBT/BCIP chromogen (Pierce), stopped by washing with tap water, air dried, and read with an ELISPOT reader (Hitech Instruments). The number of spot-forming cells (SFCs) per 10^6 PBMCs was calculated. Media backgrounds were consistently less than 15 SFCs per 10^6 PBMCs.

Received 9 October; accepted 16 November 2001.

- Barouch, D. H. *et al.* Control of viremia and prevention of clinical AIDS in rhesus monkeys by cytokine-augmented DNA vaccination. *Science* **290**, 486–492 (2000).
- Amara, R. R. *et al.* Control of a mucosal challenge and prevention of clinical AIDS in rhesus monkeys by a multiprotein DNA/MVA vaccine. *Science* **292**, 69–74 (2001).
- Rose, N. F. *et al.* An effective AIDS vaccine based on live attenuated vesicular stomatitis virus recombinants. *Cell* **106**, 539–549 (2001).
- Shiver, J. W. *et al.* Replication-incompetent adenoviral vaccine vector elicits effective anti-immunodeficiency virus immunity. *Nature* **415**, 331–335 (2002).
- Phillips, R. E. *et al.* Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition. *Nature* **354**, 453–459 (1991).
- Borrow, P. *et al.* Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus. *Nature Med.* **3**, 205–211 (1997).
- Price, D. A. *et al.* Positive selection of HIV-1 cytotoxic T lymphocyte escape variants during primary infection. *Proc. Natl Acad. Sci. USA* **94**, 1890–1895 (1997).
- Koenig, S. *et al.* Transfer of HIV-1-specific cytotoxic T lymphocytes to an AIDS patient leads to selection for mutant HIV variants and subsequent disease progression. *Nature Med.* **1**, 330–336 (1995).
- Goulder, P. J. *et al.* Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. *Nature Med.* **3**, 212–217 (1997).
- Goulder, P. J. *et al.* Evolution and transmission of stable CTL escape mutations in HIV infection. *Nature* **412**, 334–338 (2001).
- Evans, D. T. *et al.* Virus-specific cytotoxic T-lymphocyte responses select for amino-acid variation in simian immunodeficiency virus Env and Nef. *Nature Med.* **5**, 1270–1276 (1999).
- Allen, T. M. *et al.* Tat-specific cytotoxic T lymphocytes select for SIV escape variants during resolution of primary viraemia. *Nature* **407**, 386–390 (2000).
- Chen, Z. W. *et al.* Simian immunodeficiency virus evades a dominant epitope-specific cytotoxic T lymphocyte response through a mutation resulting in the accelerated dissociation of viral peptide and MHC class I. *J. Immunol.* **164**, 6474–6479 (2000).
- Reimann, K. A. *et al.* A chimeric simian/human immunodeficiency virus expressing a primary patient human immunodeficiency virus type 1 isolate *env* causes an AIDS-like disease after in vivo passage in rhesus monkeys. *J. Virol.* **70**, 6922–6928 (1996).
- Miller, M. D., Yamamoto, H., Hughes, A. H., Watkins, D. I. & Letvin, N. L. Definition of an epitope and MHC class I molecule recognized by Gag-specific cytotoxic T lymphocytes in SIVmac-infected rhesus monkeys. *J. Immunol.* **147**, 320–329 (1991).
- Allen, T. M. *et al.* Characterization of the peptide binding motif of a rhesus MHC class I molecule (Mamu-A*01) that binds an immunodominant CTL epitope from SIV. *J. Immunol.* **160**, 6062–6071 (1998).
- Altman, J. D. *et al.* Phenotypic analysis of antigen-specific T lymphocytes. *Science* **274**, 94–96 (1996).
- Kuroda, M. J. *et al.* Analysis of Gag-specific cytotoxic T lymphocytes in simian immunodeficiency virus-infected rhesus monkeys by cell staining with a tetrameric major histocompatibility complex class I-peptide complex. *J. Exp. Med.* **187**, 1373–1381 (1998).
- Egan, M. A. *et al.* Use of a major histocompatibility complex class I/peptide/B2M tetramers to quantitate CD8⁺ cytotoxic T lymphocytes specific for dominant and nondominant viral epitopes in simian–human immunodeficiency virus-infected rhesus monkeys. *J. Virol.* **73**, 5466–5472 (1999).
- Crawford, J. M. *et al.* Characterization of primary isolate-like variants of simian–human immunodeficiency virus. *J. Virol.* **73**, 10199–10207 (1999).
- Wolinsky, S. M. *et al.* Adaptive evolution of human immunodeficiency virus-type 1 during the natural course of infection. *Science* **272**, 537–542 (1996).

- Allen, T. M. *et al.* CD8⁺ lymphocytes from simian immunodeficiency virus-infected rhesus macaques recognize 14 different epitopes bound by the major histocompatibility complex class I molecule mamu-A*01: implications for vaccine design and testing. *J. Virol.* **75**, 738–749 (2001).
- Schmitz, J. E. *et al.* Control of viremia in simian immunodeficiency virus infection by CD8⁺ lymphocytes. *Science* **283**, 857–860 (1999).
- Jin, X. *et al.* Dramatic rise in plasma viremia after CD8⁺ T cell depletion in simian immunodeficiency virus-infected macaques. *J. Exp. Med.* **189**, 991–998 (1999).
- Polacino, P. *et al.* Limited breadth of the protective immunity elicited by simian immunodeficiency virus SIVmac gp160 vaccines in a combination immunization regimen. *J. Virol.* **73**, 618–630 (1999).
- Ogg, G. S. *et al.* Decay kinetics of human immunodeficiency virus-specific effector cytotoxic T lymphocytes after combination antiretroviral therapy. *J. Virol.* **73**, 797–800 (1999).
- Kalams, S. A. *et al.* Levels of human immunodeficiency virus type 1-specific cytotoxic T-lymphocyte effector and memory responses decline after suppression of viremia with highly active antiretroviral therapy. *J. Virol.* **73**, 6721–6728 (1999).

Acknowledgements

We are grateful to T. Fu, J. Shiver, K. Reimann, C. Lord, A. Miura, R. Khunkhun, W. Wagner, S. Tyeryar, C. Crabbs and J. Valley-Ogunro for advice, assistance and reagents.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.H.B. (e-mail: dan_barouch@hotmail.com).

Leptin stimulates fatty-acid oxidation by activating AMP-activated protein kinase

Yasuhiko Minokoshi*, Young-Bum Kim*, Odile D. Peroni*, Lee G. D. Fryer†, Corinna Müller*, David Carling† & Barbara B. Kahn*

* Division of Endocrinology, Diabetes and Metabolism, Beth Israel Deaconess Medical Center and Department of Medicine, Harvard Medical School, Boston, Massachusetts 02215, USA

† The Cellular Stress Group, MRC Clinical Sciences Centre, Imperial College School of Medicine, Hammersmith Hospital, London W12 0NN, UK

Leptin is a hormone secreted by adipocytes that plays a pivotal role in regulating food intake, energy expenditure and neuroendocrine function¹. Leptin stimulates the oxidation of fatty acids² and the uptake of glucose^{3,4}, and prevents the accumulation of lipids in nonadipose tissues, which can lead to functional impairments known as “lipotoxicity”⁵. The signalling pathways that mediate the metabolic effects of leptin remain undefined. The 5'-AMP-activated protein kinase (AMPK) potently stimulates fatty-acid oxidation in muscle by inhibiting the activity of acetyl coenzyme A carboxylase (ACC)^{6,7}. AMPK is a heterotrimeric enzyme that is conserved from yeast to humans and functions as a ‘fuel gauge’ to monitor the status of cellular energy⁶. Here we show that leptin selectively stimulates phosphorylation and activation of the $\alpha 2$ catalytic subunit of AMPK ($\alpha 2$ AMPK) in skeletal muscle, thus establishing a previously unknown signalling pathway for leptin. Early activation of AMPK occurs by leptin acting directly on muscle, whereas later activation depends on leptin functioning through the hypothalamic-sympathetic nervous system axis. In parallel with its activation of AMPK, leptin suppresses the activity of ACC, thereby stimulating the oxidation of fatty acids in muscle. Blocking AMPK activation inhibits the phosphorylation of ACC stimulated by leptin. Our data identify AMPK as a principal mediator of the effects of leptin on fatty-acid metabolism in muscle.

Impaired fuel metabolism is an early pathogenic factor in obesity and type 2 diabetes⁸. Skeletal muscle is a principal site of glucose and fatty-acid use, and is one of the primary tissues responsible for

insulin resistance in obesity and type 2 diabetes⁸. Increased lipid stores in nonadipose tissues such as muscle are linked to functional impairments, called lipotoxicity⁵, which lead to insulin resistance⁹ and impaired insulin secretion⁵. Treatment with leptin *in vivo* markedly improves insulin sensitivity⁸, possibly owing, in part, to a depletion of triglyceride stores in skeletal muscle^{5,10}. Although many of the effects of leptin are exerted in the hypothalamus¹¹, leptin stimulates fatty-acid oxidation and reduces fatty-acid esterification directly in muscle *in vitro*².

ACC provides a pivotal step in fuel metabolism as it links fatty-acid and carbohydrate metabolism through the shared intermediate acetyl coenzyme A (CoA)¹². The β -isoform of ACC (ACC- β), which is the predominant isoform in skeletal muscle, is an essential regulator of fatty-acid oxidation¹². ACC- β knockout mice show increased oxidation of fatty acids in muscle and are lean¹². Phosphorylation of ACC- β by AMPK leads to inhibition of ACC activity, a fall in malonyl-CoA content, and a subsequent increase in fatty-acid oxidation by disinhibiting carnitine palmitoyltransferase 1 in skeletal muscle^{6,7,9}. Thus, we thought that AMPK, through inhibition of ACC, might mediate the effects of leptin on lipid metabolism in muscle.

An intrahypothalamic (i.h.p.) injection of leptin (1 ng) in free-moving FVB mice increased the activity of $\alpha 2$ AMPK in soleus muscle (Fig. 1a). The effect peaked with a threefold stimulation at

1 h and was sustained for up to 6 h. In contrast, intravenous (i.v.) leptin (1 mg per kg body weight) produced a biphasic response in $\alpha 2$ AMPK activity, with a twofold rise at about 15 min, a return to baseline by 60 min, and a second twofold elevation by 6 h (Fig. 1b). $\alpha 1$ AMPK activity did not change in soleus muscle after i.v. or i.h.p. injection of leptin (data not shown), although it increased twofold after i.v. injection of AICAR (5-aminoimidazole-4-carboxamide 1- β -D-ribofuranoside, 1 g per kg), a cell-permeable activator of AMPK that is converted to ZMP^{6,7}, an AMP analogue that mimics the effects of AMP on phosphorylation and allosteric activation of AMPK.

A lower dose of leptin (50 μ g per kg, i.v.), which is closer to the physiological range (Fig. 1d), also stimulated AMPK activity in muscle at both 15 min and 6 h after injection (Fig. 1c). Notably, db/db mice, which lack the long-form leptin receptor, Ob-Rb, did not show any stimulation of AMPK, even with high dose leptin (1 mg per kg). The effect of i.v. leptin on the activation of $\alpha 2$ AMPK was more pronounced in red (slow twitch, oxidative) skeletal muscles, which have higher rates of fatty-acid oxidation, than in white (fast twitch, glycolytic) muscles (Fig. 1e), although there was a tendency for an increase in activation in white gastrocnemius muscle at 15 min after i.v. leptin ($P = 0.052$). Similarly, at 1 h after i.h.p. injection, leptin stimulated $\alpha 2$ AMPK in soleus ($292 \pm 17\%$, $P < 0.01$) and red gastrocnemius ($150 \pm 8\%$, $P < 0.05$) muscle. Serum concentrations of leptin were increased 15 min after both doses of i.v. leptin, but returned to baseline by 6 h (Fig. 1d).

To determine whether the effects of leptin on AMPK activation require the hypothalamic-sympathetic nervous system axis, we used pharmacological adrenergic blockade and two kinds of surgical denervation of hindlimb. Denervation of the sciatic nerve, which primarily impairs motor innervation, decreased the total catecholamine content to 62% of the contralateral, intact soleus muscle. Denervation of the sciatic, femoral and obturator nerves, which together blocks both sympathetic and motor innervation, decreased catecholamine to 6% of the control content (Fig. 2a).

Denervation of all three nerves blocked the ability of i.h.p. leptin or i.v. leptin 6 h after injection to stimulate $\alpha 2$ AMPK in soleus muscle (Fig. 2b). But $\alpha 2$ AMPK activation at 15 min after i.v. leptin, and after i.v. AICAR, remained intact in denervated muscle. By contrast, denervation of the sciatic nerve alone did not block leptin-induced activation of AMPK in the soleus, despite a complete loss of motor function (Fig. 2c). Thus, leptin's effect on AMPK is independent of motor nerve activity. These findings indicate that AMPK activation both at 6 h after i.v. leptin and after i.h.p. leptin may be mediated by the sympathetic nerves that innervate muscle. In contrast, activation at 15 min after i.v. leptin does not involve the sympathetic nervous system and may be direct (Fig. 2b). We incubated soleus muscle *ex vivo* with leptin to show conclusively that leptin directly stimulates AMPK activity. Leptin at 1 and 10 nM increased $\alpha 2$ AMPK activity 2.0- and 2.8-fold at 30 min, respectively (Fig. 2e).

To elucidate the pathways involved in the neurally mediated effects, we examined the impact of the α -adrenergic antagonist phentolamine on leptin-induced activation of $\alpha 2$ AMPK in soleus muscle (Fig. 2d). Consistent with the results of surgical denervation, phentolamine *in vivo* abolished AMPK activation 6 h after i.v. leptin

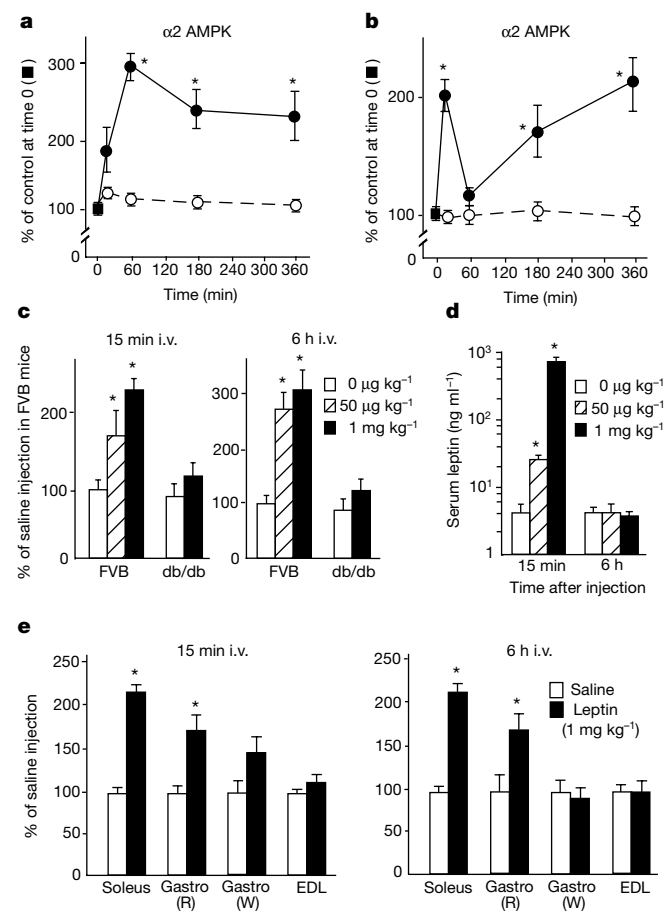


Figure 1 Leptin activates $\alpha 2$ AMPK in skeletal muscle. **a**, **b**, $\alpha 2$ AMPK activity in soleus muscle after i.h.p. (1 ng) injection (**a**) or i.v. (1 mg per kg) injection (**b**) of leptin (filled circles) or saline (open circles) ($n = 7-10$). **c**, Leptin increases $\alpha 2$ AMPK activity in soleus muscle of FVB mice, but not db/db mice, at 15 min and 6 h after i.v. injection ($n = 5-7$). **d**, Concentration of serum leptin after i.v. injection of leptin ($n = 7$). **e**, Effects of i.v. leptin on $\alpha 2$ AMPK activity in soleus, red (R) and white (W) gastrocnemius (Gastro) muscle and EDL ($n = 7$). Asterisk, $P < 0.05$ compared with saline injection.

Table 1 Adenine nucleotides in soleus muscles after injecting leptin *in vivo*

	AMP	ADP	ATP	Total
Control	0.63 $\pm$ 0.07	4.82 $\pm$ 0.25	34.42 $\pm$ 0.82	39.88 $\pm$ 0.89
Leptin i.h.p. (1 h)	0.89 $\pm$ 0.27	4.86 $\pm$ 0.31	34.40 $\pm$ 0.60	40.16 $\pm$ 0.40
Leptin i.v. (15 min)	1.32 $\pm$ 0.21*	5.48 $\pm$ 0.38	31.99 $\pm$ 1.66	38.79 $\pm$ 1.42
Leptin i.v. (6 h)	0.86 $\pm$ 0.10	4.98 $\pm$ 0.44	33.55 $\pm$ 0.62	39.35 $\pm$ 0.63

Data represent the mean $\pm$ s.e.m ($n = 6$) in nmol per muscle. i.h.p., intrahypothalamic injection; i.v., intravenous injection (1 mg per kg body weight).

* $P < 0.05$ compared with control value.

and after i.h.p. leptin, but not the early activation by i.v. leptin or the activation by AICAR. Thus, the early effect of i.v. leptin may be direct, but the later effect depends on the α -adrenergic nervous system. This was confirmed by *ex vivo* studies. Soleus muscle incubated *ex vivo* with phenylephrine, an α -adrenergic agonist, showed stimulation of AMPK activity that equalled that of leptin (Fig. 2e). In contrast, the β -adrenergic agonist, isoproterenol, had no effect. Leptin had no effect on AMPK activity in soleus muscle from db/db mice, whereas phenylephrine stimulated AMPK two-

fold (Fig. 2e). Thus, the α -adrenergic, but not the β -adrenergic, pathway is involved in the sympathetic nerve-mediated effect of leptin on AMPK, and these effects are mediated through Ob-Rb.

Leptin administration *in vivo* increased phosphorylation of Thr 172 in the α -subunit of AMPK (Fig. 2f). Phosphorylation of this residue is essential for activation of AMPK¹³. Denervation did not block phosphorylation at 15 min but blocked it at 6 h, indicating that the former effect is direct, whereas the latter requires the hypothalamic-sympathetic nervous system.

Known activators of AMPK include increased concentrations of cellular AMP^{6,7}, decreased cellular pH (ref. 14) and an increased creatine/phosphocreatine ratio¹⁴. AMP-independent activation of AMPK has also been observed (ref. 15; and D.C., unpublished data). The AMP content of soleus muscle increased about twofold at 15 min after i.v. leptin, but not at 6 h after i.v. leptin or after i.h.p. leptin (Table 1). Neither ATP nor ADP changed in response to leptin. Thus, AMPK activation at 15 min after i.v. leptin may result from increased concentrations of cellular AMP, but the activation mediated by the sympathetic nervous system does not involve changes in AMP, ADP or ATP.

To determine whether AMPK activation mediates leptin's effects

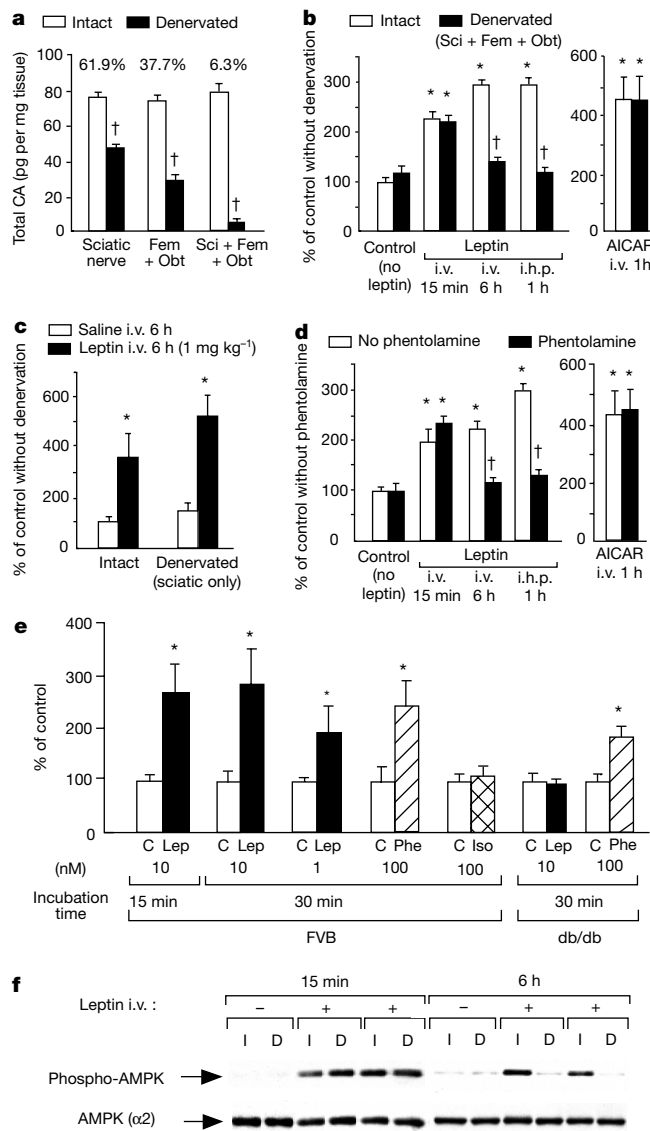


Figure 2 Leptin activates α 2 AMPK in soleus muscle through two different mechanisms. **a**, Effect of surgical denervation of the sciatic (Sci), femoral (Fem) and obturator (Obt) nerves on total catecholamine (CA) content in soleus. **b**, Effects of surgical denervation of the sciatic, femoral and obturator nerves on leptin (1 mg per kg i.v. or 1 ng i.h.p.) or AICAR-stimulated α 2 AMPK activity in soleus. **c**, Surgical denervation of the sciatic nerve alone does not inhibit leptin-induced α 2 AMPK activation in soleus. **d**, Effects of α -adrenergic antagonist, phentolamine, *in vivo* on leptin (1 mg per kg i.v. or 1 ng i.h.p.) or AICAR-stimulated α 2 AMPK activity in soleus. (**b–d**, $n = 6$). Asterisk, $P < 0.01$ compared with control without leptin, $P < 0.01$ compared with soleus without denervation or phentolamine. **e**, Effects of leptin (Lep), phenylephrine (Phe) or isoproterenol (Iso) on α 2 AMPK activity in soleus *ex vivo*. Asterisk, $P < 0.05$ compared with control (C) ($n = 5–10$). **f**, Phosphorylation of α AMPK and total protein level of α 2 AMPK in soleus 15 min or 6 h after i.v. leptin (1 mg per kg). I, intact; D, denervated (sciatic, femoral and obturator nerves).

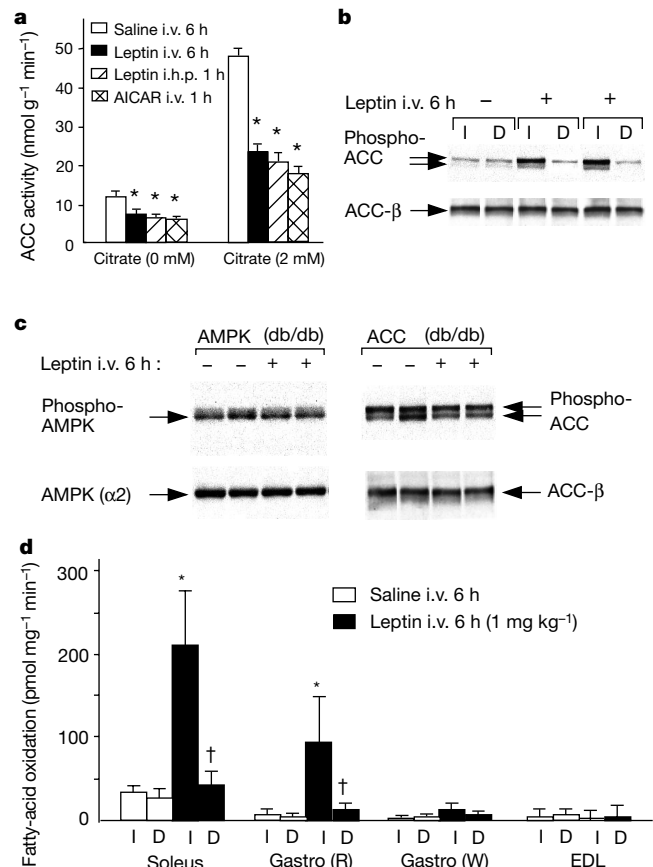


Figure 3 Leptin inhibits acetyl-CoA carboxylase (ACC) activity and activates fatty-acid oxidation in red muscle *in vivo*. Leptin fails to phosphorylate AMPK or ACC in db/db muscle. **a**, *In vivo*, (1 mg per kg) or i.h.p. leptin, or AICAR, decreases ACC activity in soleus muscle ($n = 7$). **b**, Phosphorylation and protein levels of ACC in soleus 6 h after i.v. leptin (1 mg per kg). **c**, Phosphorylation of α AMPK and ACC and total protein level of α 2 AMPK and ACC in soleus muscle of db/db mice 6 h after i.v. leptin (1 mg per kg). **d**, *In vivo* fatty-acid oxidation in muscles in FVB mice 6 h after i.v. leptin (1 mg per kg). I, intact; D, denervated (sciatic, femoral and obturator nerves). Serum fatty-acid concentration 6 h after i.v. leptin or saline was $459 \pm 65 \mu\text{M}$ and $497 \pm 61 \mu\text{M}$, respectively. Asterisk, $P < 0.05$ compared with saline injection. Dagger, $P < 0.05$ versus intact muscle from the same animal.

on fatty-acid metabolism, we examined the activity of ACC^{6,7,9}. I.v. and i.h.p. leptin suppressed ACC activity in soleus muscle by about 50%, which is comparable to the inhibition by AICAR (Fig. 3a). These effects are probably caused by phosphorylation of ACC by AMPK, as ACC phosphorylation in soleus muscle increased after i.v. leptin, whereas concentrations of ACC protein did not change (Fig. 3b). ACC phosphorylation 6 h after i.v. leptin was blocked by denervation of the sciatic, femoral and obturator nerves. Levels of ACC phosphorylation strongly correlated with AMPK phosphorylation ($R = 0.94 \pm 0.10$, $P < 0.01$). In db/db mice, leptin did not stimulate phosphorylation of AMPK or ACC (Fig. 3c), indicating that these effects of leptin require Ob-Rb.

I.v. leptin increased fatty-acid utilization in soleus and red gastrocnemius muscle (3.2- and 3.9-fold, respectively, $P < 0.01$) at 6 h, but not in white gastrocnemius muscle or extensor digitorum longus (EDL) (data not shown). Fatty-acid incorporation into storage lipids was not increased by leptin (data not shown); however, leptin increased the rate of fatty-acid oxidation 6.4-fold and 15.3-fold in soleus and red gastrocnemius muscle, respectively (Fig. 3d). This effect was abolished by denervation of the sciatic, femoral and obturator nerves (Fig. 3d) but not by denervation of the sciatic nerve alone (data not shown). Thus, leptin stimulates fatty-acid oxidation in red muscle, and the effect at 6 h is mediated through the sympathetic nervous system.

To determine whether the AMPK pathway is required for the effects of leptin on ACC phosphorylation, and thus fatty-acid metabolism, we blocked AMPK activation using a catalytically inactive α -subunit of AMPK that has a dominant-negative effect on AMPK activity¹⁶. Dominant-negative AMPK inhibited basal AMPK activity in H-2K^b muscle cells^{17,18} by 63%. Ob-Rb has been detected in H-2K^b cells (P. Shepherd, personal communication), and these cells are highly responsive to insulin and AICAR¹⁸. Leptin induced AMPK activity 1.8-fold in control cells. This effect was completely blocked by dominant-negative AMPK (Fig. 4a), indicating a full dominant-negative effect on leptin signalling through AMPK. Dominant-negative AMPK inhibited basal ACC phosphorylation by 35% (Fig. 4b). Leptin induced ACC phosphorylation 1.9-fold, and this effect was also blocked by dominant-negative AMPK. Thus, activation of AMPK is necessary for leptin to inhibit ACC.

Our data show that leptin activates $\alpha 2$ AMPK in red muscle through two distinct mechanisms: one involves the hypothalamus and the α -adrenergic pathway; the other is independent of the sympathetic nervous system and appears to be direct. In support of the first mechanism, peripheral leptin injection increases sym-

thetic nerve activity that innervates hindlimbs¹⁹, with a time course that is consistent with the late activation of $\alpha 2$ AMPK in muscle after i.v. leptin. Furthermore, AMPK can be activated by the stimulation of G_q-coupled receptors, including α -adrenergic receptors²⁰. α -adrenergic receptors have important functions in muscle fibres^{21–23}. These results support our conclusions that the delayed effects of peripheral leptin are through hypothalamic neurons that stimulate the α -adrenergic sympathetic pathway to activate $\alpha 2$ AMPK in muscle. In contrast, the early effect is independent of sympathetic and motor nerves, can be induced directly in muscle incubated with leptin *ex vivo*, and involves a transient increase in AMP concentrations.

Activation of AMPK by leptin results in phosphorylation and inhibition of ACC. The potential for a major impact on metabolism is illustrated by the markedly increased oxidation of fatty acids and leanness in ACC knockout mice¹². Leptin potentially stimulates fatty-acid oxidation in muscle *in vivo*. This appears to be mediated by AMPK, as dominant-negative AMPK blocks leptin-stimulated phosphorylation of ACC. A physiological role of the hyperleptinemia of caloric excess may be to protect nonadipocytes from steatosis and lipotoxicity by preventing the upregulation of lipogenesis and by increasing the oxidation of fatty acids⁵. Therefore, the AMPK pathway might provide targets for therapeutic agents to reduce lipotoxicity in obesity and type 2 diabetes. □

Methods

Animals

Male FVB and db/db mice (aged 8–10 weeks) were purchased from Taconic and Jackson Labs, respectively. Mice were housed in individual cages in a temperature-controlled environment with a 12-h light/dark cycle and given chow (Formulab 5008; Farmer's Exchange) and water freely. Five days before the experiment, we implanted chronic cannulas into the medial hypothalamus or jugular vein, and performed unilateral surgical denervation of hindlimb by severing the femoral, obturator and/or sciatic nerves.

Injection of leptin, AICAR or saline was started at 9:00 to 10:00 a.m. (lights on 6:00). Food and water were available during the experiments. Recombinant murine leptin (1 ng) (Eli Lilly) dissolved in 0.2 μ l of saline solution was injected into the medial hypothalamus in conscious unrestrained mice through the implanted brain cannula. In another group of mice, 1 mg or 50 μ g per kg (body weight) of leptin or 1 g per kg (body weight) of AICAR was injected intravenously in conscious unrestrained mice through the jugular vein catheter. We injected control mice with the same volume of saline either i.v. or into the medial hypothalamus. After the mice were killed with ketamine and xylazine solution, at each time point the skeletal muscles were rapidly dissected and frozen in liquid nitrogen. Mice without treatments were used as controls at time 0. Where specified, the α -adrenergic receptor antagonist phenolamine (10 mg per kg; Sigma) was injected intraperitoneally (i.p.), 30 min before and 3 h after injection of leptin or saline. The Institutional Animal Care and Use Committee (Beth Israel Deaconess Medical Center) approved all studies.

Measurement of enzyme activities

To measure isoform-specific AMPK activity in skeletal muscles, we immunoprecipitated AMPK from muscle lysates (100 μ g of protein) with specific antibodies²⁴ against the $\alpha 1$ or $\alpha 2$ catalytic subunits bound to protein-G/sepharose beads. In the study with H-2K^b cells, AMPK was immunoprecipitated with an antibody¹⁵ against the β -regulatory subunit. We measured kinase activity using synthetic 'SAMS' peptide and [γ -³²P]ATP²⁵.

The activity of ACC in muscle lysates was measured by ¹⁴C₂ fixation to acid-stable products in the presence or absence of citrate (2 mM), an allosteric activator of ACC²⁶.

Measurement of fatty-acid utilization in muscles *in vivo*

Total fatty-acid utilization (R_t) and incorporation of fatty acids into storage lipids (R'_s) in individual muscles were determined in conscious mice, using [9,10-³H]-(R)-2-bromopalmitate (AstraZeneca) and [U-¹⁴C]palmitate (Amersham Pharmacia)²⁷. The rate of fatty-acid oxidation in individual muscles was calculated by subtracting R'_s from R_t . We determined the lumped constant for calculating R_t by the ratio of R_t to R'_s in etomoxir-treated mice (20 mg per kg, i.p.)²⁷. The fatty-acid concentration in serum was measured by a colorimetric method (NEFA C, Wako).

In vitro studies with isolated soleus muscle

Mice were anaesthetized with ketamine and xylazine solution, and soleus muscles were dissected and incubated with Krebs–Henseleit bicarbonate buffer supplemented with 5 mM HEPES, 5 mM glucose and 2 mM pyruvate, and gassed with O₂/CO₂ (95:5). After 30 min of pre-incubation, leptin (1 or 10 nM), L-phenylephrine (100 nM; Sigma) or ($\pm$)-isoproterenol (100 nM; Sigma) was added for an additional 15 or 30 min. We measured $\alpha 2$ AMPK activity as described above. The contralateral soleus muscle was incubated without leptin and used as a control.

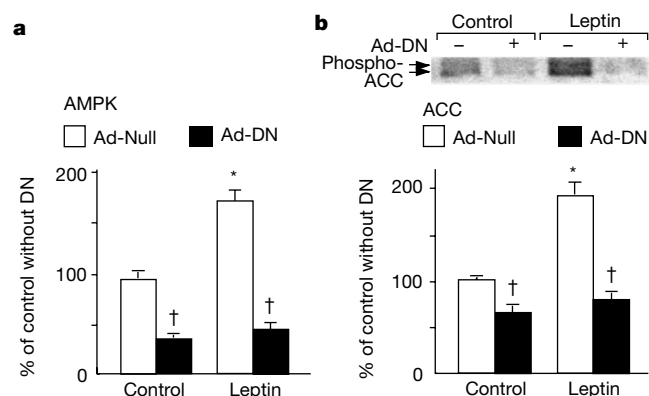


Figure 4 Dominant-negative AMPK inhibits leptin-induced AMPK activation and ACC phosphorylation in H-2K^b cells. Effects of dominant-negative AMPK on AMPK activity ($n = 5–7$) (a) and ACC phosphorylation ($n = 6$) (b) in H-2K^b cells. Leptin (100 nM) was present for 2 h. Ad-Null, cells infected with control vector; Ad-DN, dominant-negative AMPK-infected cells. Asterisk, $P < 0.01$ compared with control without leptin. Dagger, $P < 0.05$ compared with Ad-Null.

Studies with H-2K^b cells

H-2K^b cells derived from the muscle of heterozygous H-2K^b-ts-A58 transgenic mice¹⁷ were differentiated¹⁸ and transfected with adenovirus containing control vector or dominant-negative AMPK¹⁶. Five days later, cells were treated with leptin (100 nM) for 2 h in HEPES-buffered saline containing 5 mM glucose, and the AMPK activity and phosphorylation of ACC were examined.

Measurement of catecholamine and AMP, ADP, ATP content

Soleus muscles were homogenized with 0.5 N perchloric acid and centrifuged. We measured total catecholamine content in the supernatant by a radioenzymatic assay kit (Amersham Pharmacia). AMP, ADP and ATP were determined by anion exchange chromatography using a SMART system (Amersham Pharmacia)²⁸.

Western blot analysis

Phosphorylation of the α -subunit of AMPK and ACC in soleus lysates (40 μ g of protein) was determined with 4–15% gradient SDS acrylamide gels using antibodies against phosphopeptides based on the amino-acid sequence surrounding Thr 172 of the α -subunit of human AMPK (Cell Signaling) and Ser 79 of rat ACC (Upstate Biotech.). We determined protein levels of α 2 AMPK and ACC by using a specific antibody for the α 2 subunit of AMPK (ref. 24) and streptavidin–horseradish peroxidase (Amersham Pharmacia)²⁹, respectively.

Statistical analysis

All values are the mean $\pm$ s.e.m. Data were evaluated by factorial analysis of variance and Newman–Keuls multiple range test. The difference was considered to be significant if $P < 0.05$.

Received 28 August; accepted 12 October 2001.

- Friedman, J. M. & Halaas, J. L. Leptin and the regulation of body weight in mammals. *Nature* **395**, 763–770 (1998).
- Muio, D. M. *et al.* Leptin directly alters lipid partitioning in skeletal muscle. *Diabetes* **46**, 1360–1363 (1997).
- Kamohara, S., Surcelin, R., Halaas, J. L., Friedman, J. M. & Charron, M. J. Acute stimulation of glucose metabolism in mice by leptin treatment. *Nature* **389**, 374–377 (1997).
- Minokoshi, Y., Haque, M. S. & Shimazu, T. Microinjection of leptin into the ventromedial hypothalamus increases glucose uptake in peripheral tissues in rats. *Diabetes* **48**, 287–291 (1999).
- Unger, R. H., Zhou, Y.-T. & Orci, L. Regulation of fatty acid homeostasis in cells: novel role of leptin. *Proc. Natl Acad. Sci. USA* **96**, 2327–2332 (1999).
- Hardie, D. G., Carling, D. & Carlson, M. The AMP-activated/SNF1 protein kinase subfamily: metabolic sensors of the eukaryotic cell? *Ann. Rev. Biochem.* **67**, 821–855 (1998).
- Winder, W. W. & Hardie, D. G. AMP-activated protein kinase, a metabolic master switch: possible roles in type 2 diabetes. *Am. J. Physiol.* **277**, E1–E10 (1999).
- Kahn, B. B. & Flier, J. S. Obesity and insulin resistance. *J. Clin. Invest.* **106**, 473–481 (2000).
- Ruderman, N. B., Saha, A. K., Vavvas, D. & Witters, L. A. Malonyl-CoA, fuel sensing, and insulin resistance. *Am. J. Physiol.* **276**, E1–E18 (1999).
- Lee, Y. *et al.* Liporegulation in diet-induced obesity. The antisteatotic role of hyperleptinemia. *J. Biol. Chem.* **276**, 5629–5635 (2001).
- Schwartz, M. W., Woods, S. C., Porte, D. Jr, Seeley, R. J. & Baskin, D. G. Central nervous system control of food intake. *Nature* **404**, 661–671 (2000).
- Abu-Elheiga, L., Matzuk, M. M., Abo-Hashema, K. A. H. & Wakil, S. J. Continuous fatty acid oxidation and reduced fat storage in mice lacking acetyl-CoA carboxylase 2. *Science* **291**, 2613–2616 (2001).
- Stein, S. C., Woods, A., Jones, N. A., Davison, M. D. & Carling, D. The regulation of AMP-activated protein kinase by phosphorylation. *Biochem. J.* **345**, 437–443 (2000).
- Ponticos, M. *et al.* Dual regulation of the AMP-activated protein kinase provides a novel mechanism for the control of creatine kinase in skeletal muscle. *EMBO J.* **17**, 1688–1699 (1998).
- Higaki, Y., Hirshman, M. F., Fujii, N. & Goodyear, L. J. Nitric oxide increases glucose uptake through a mechanism that is distinct from the insulin and contraction pathways in rat skeletal muscle. *Diabetes* **50**, 241–247 (2001).
- Woods, A. *et al.* Characterization of the role of AMP-activated protein kinase in the regulation of glucose-activated gene expression using constitutively active and dominant negative forms of the kinase. *Mol. Cell. Biol.* **20**, 6704–6711 (2000).
- Jat, P. S. *et al.* Direct derivation of conditionally immortal cell lines from an H-2K^b-ts-A58 transgenic mouse. *Proc. Natl Acad. Sci. USA* **88**, 5096–5100 (1991).
- Fryer, L. G. D. *et al.* Activation of glucose transport by AMP-activated protein kinase via stimulation of nitric oxide synthase. *Diabetes* **49**, 1978–1985 (1999).
- Haynes, W. G., Morgan, D. A., Walsh, S. A., Mark, A. L. & Sivitz, W. I. Receptor-mediated regional sympathetic nerve activation by leptin. *J. Clin. Invest.* **100**, 270–278 (1997).
- Kishi, K. *et al.* AMP-activated protein kinase is activated by the stimulations of G_q-coupled receptors. *Biochem. Biophys. Res. Commun.* **276**, 16–22 (2000).
- Martin, W. H., Tolley, T. K. & Saffitz, J. E. Autoradiographic delineation of skeletal muscle α -adrenergic receptor distribution. *Am. J. Physiol.* **259**, H1402–H1408 (1990).
- Akaike, N. Sodium pump in skeletal muscle: central nervous system-induced suppression by α -adrenoreceptors. *Science* **213**, 1252–1254 (1981).
- Stafford I. L. & Jacobs, B. L. Noradrenergic modulation of the masseteric reflex in behaving cats. I. Pharmacological studies. *J. Neurosci.* **10**, 91–98 (1990).
- Woods, A., Salt, I., Scott, J., Hardie, D. G. & Carling, D. The α 1 and α 2 isoforms of the AMP-activated protein kinase have similar activities in rat liver but exhibit differences in substrate specificity *in vitro*. *FEBS Lett.* **397**, 347–351 (1996).
- Hayashi, T. *et al.* Metabolic stress and altered glucose transport. Activation of AMP-activated protein kinase as a unifying coupling mechanism. *Diabetes* **49**, 527–531 (2000).

- Goodwin, G. W. & Taegtmeyer, H. Regulation of fatty acid oxidation of the heart by MCD and ACC during contractile stimulation. *Am. J. Physiol.* **277**, E772–E777 (1999).
- Oakes, N. D. *et al.* Development and initial evaluation of a novel method for assessing tissue-specific plasma free fatty acid utilization *in vivo* using (R)-2-bromopalmitate tracer. *J. Lipid. Res.* **40**, 1155–1169 (1999).
- Corton, J. M., Gillespie, J. G. & Hardie, D. G. Role of the AMP-activated protein kinase in the cellular stress response. *Curr. Biol.* **4**, 315–324 (1994).
- Vavvas, D. *et al.* Contraction-induced changes in acetyl-CoA carboxylase and 5'-AMP-activated kinase in skeletal muscle. *J. Biol. Chem.* **272**, 13255–13261 (1997).

Acknowledgements

We thank N. D. Oakes for the gift of [9,10-³H]-(R)-2-bromopalmitate, and B. B. Lowell and L. A. Witters for helpful advice. This work was supported by NIH grants (B.B.K.), the Boston Obesity Nutrition Research Center (Y.-B.K.), and The Kato Memorial Trust for Nambyo Research and Mitsui Life Social Welfare Foundation (Y.M.).

Correspondence and requests for materials should be addressed to B.B.K. (e-mail: bkahn@caregroup.harvard.edu).

Energetic landscape of α -lytic protease optimizes longevity through kinetic stability

Sheila S. Jaswal*, Julie L. Sohl, Jonathan H. Davis* & David A. Agard

Howard Hughes Medical Institute and the Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, California 94143-0448, USA

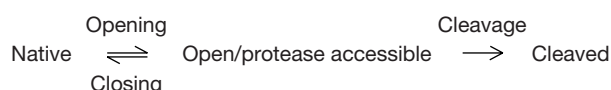
During the evolution of proteins the pressure to optimize biological activity is moderated by a need for efficient folding. For most proteins, this is accomplished through spontaneous folding to a thermodynamically stable and active native state. However, in the extracellular bacterial α -lytic protease (α LP) these two processes have become decoupled. The native state of α LP is thermodynamically unstable, and when denatured, requires millennia ($t_{1/2} \sim 1,800$ years)¹ to refold. Folding is made possible by an attached folding catalyst, the pro-region, which is degraded on completion of folding, leaving α LP trapped in its native state by a large kinetic unfolding barrier ($t_{1/2} \sim 1.2$ years)¹. α LP faces two very different folding landscapes: one in the presence of the pro-region controlling folding, and one in its absence restricting unfolding. Here we demonstrate that this separation of folding and unfolding pathways has removed constraints placed on the folding of thermodynamically stable proteins, and allowed the evolution of a native state having markedly reduced dynamic fluctuations. This, in turn, has led to a significant extension of the functional lifetime of α LP by the optimal suppression of proteolytic sensitivity.

α LP and virtually all other extracellular bacterial proteases are synthesized with a pro-region and where examined, the pro-regions are required for folding of the attached protease². Thus, many of these other proteases are expected to share the unusual features of folding energetics of α LP. The biological function of secreted bacterial proteases is to provide nutrients for the bacterium. As they function in the soil, there is no need for their regulated turnover. Therefore the native states of these proteases are likely to be highly optimized for survival under harsh, highly proteolytic conditions. We have investigated the possibility that α LP—and

* Present addresses: Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520-8114, USA (S.S.J.); Lexigen Pharmaceuticals, 125 Hartwell Avenue, Lexington, Massachusetts 02421, USA (J.H.D.).

probably other pro-proteases and proteins functioning under extreme conditions—has evolved kinetic stability via unique native-state properties in addition to the large unfolding barrier¹ (Fig. 1), as a means to achieve optimal proteolytic resistance, thereby extending biological function.

This proposal raises the question of how native-state properties arising from kinetic stability lead to effective protease resistance. As between four and eight residues of the substrate peptide chain are required to fit in a precise manner into the deep cleft that forms the protease active site³, flexible regions of a protein are the most accessible targets for proteolysis⁴. Even proteins lacking proteolytically accessible flexible loops in their native states are conformationally dynamic, constantly sampling open, proteolytically vulnerable conformations in processes ranging from small breathing motions to global unfolding: a finding that has been well established from native-state hydrogen exchange studies^{5,6}. Thus, proteolytic degradation can be described by the following conformational equilibrium reaction^{7,8}:



Therefore, to optimally avoid proteolytic attack and prolong its lifetime, a protein must not only eliminate native-state proteolysis but also restrict those unfolding and dynamic opening events leading to protease accessibility.

To test whether α LP shows restricted native-state dynamics in addition to its severely restricted global unfolding¹, we assessed the ability of water to penetrate the protease by hydrogen–deuterium exchange (H–X) monitored by two-dimensional nuclear magnetic resonance (NMR) spectroscopy. H–X is a sensitive and residue-specific probe of conformational fluctuations based on the rate at which amide protons exchange with solvent deuterons⁶. Amide protons buried in the core of a protein are protected from solvent exchange. The degree of protection from exchange is quantified using protection factors (P_f) corresponding to the ratio of the intrinsic exchange rate (k_{int}) of the solvent-exposed amide to the

observed exchange rate (k_{obs}): $P_f = k_{\text{int}}/k_{\text{obs}}$. Protons on the surface or in dynamic regions of the protein sampling partially unfolded conformations exchange much more rapidly and have lower P_f values than those in rigid regions of the protein.

As shown in Fig. 2, α LP displays extremely high protection from H–X across the entire molecule. Although most proteins typically display a ‘slow exchange core’⁹ of only 3–8 residues with their highest P_f values ranging from 10^4 to 10^8 , more than half of the amide protons of α LP (103) have P_f values greater than 10^4 . Notably, 31 of those amide protons display P_f values greater than 10^9 ; 18 are so highly protected that no exchange was measured at all, even after six months at pH 9, corresponding to P_f values of $>10^{9.4}$ to $10^{10.7}$. Not only is this degree of protection among the highest to be measured in a protein¹⁰, but the most slowly exchanging residues are spread throughout both domains of the protein instead of being localized to one discrete ‘core’, as is usually seen. Therefore α LP has a conformational rigidity well beyond that seen for traditional, thermodynamically stabilized proteins.

The high degree of rigidity observed by H–X extends previous thermodynamic measurements demonstrating that the native state of α LP is enthalpically favoured over its partially unfolded intermediate by 18 kcal mol^{−1} (ref. 1), and as a consequence, its thermodynamic instability must derive from an excessive loss in entropy during the course of folding. A possible source for this unusually large entropy decrease is the high number of glycines found in the sequence of α LP and all other pro-region-containing members of the chymotrypsin family (18% on average compared with 9% in

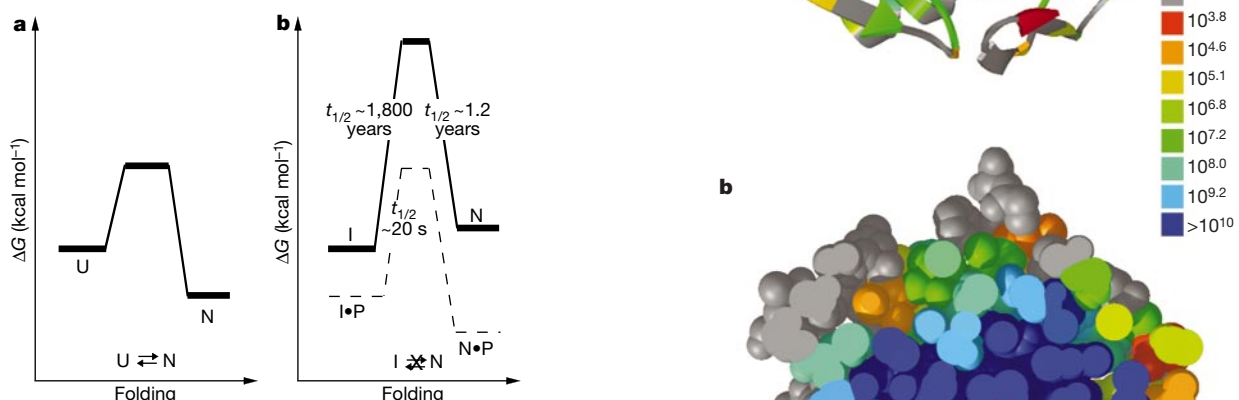


Figure 1 Folding free-energy diagrams for thermodynamic compared with kinetic stability. ΔG , free energy difference. **a**, A typical protein that folds on its own to a thermodynamically stable native state (N) is in equilibrium with its unfolded states (U), therefore it is constantly sampling protease-accessible conformations. **b**, In the absence of the pro-region (solid line) the large barrier prevents the native state of α LP (N) from being in equilibrium with the more stable intermediate (I) and other unfolded conformations. In the presence of the pro-region (P; dashed line) the barrier is lowered and N is stabilized. Refolding rates were determined by measuring an increase in α LP activity owing to formation of N over time during incubation of I in the absence or presence of pro¹. The unfolding rate was determined from fluorescence measurements of inactive S195A α LP unfolded in varying amounts of denaturant and extrapolated to water¹. Experiments were performed at 4 °C, pH 5.0 (ref. 1).

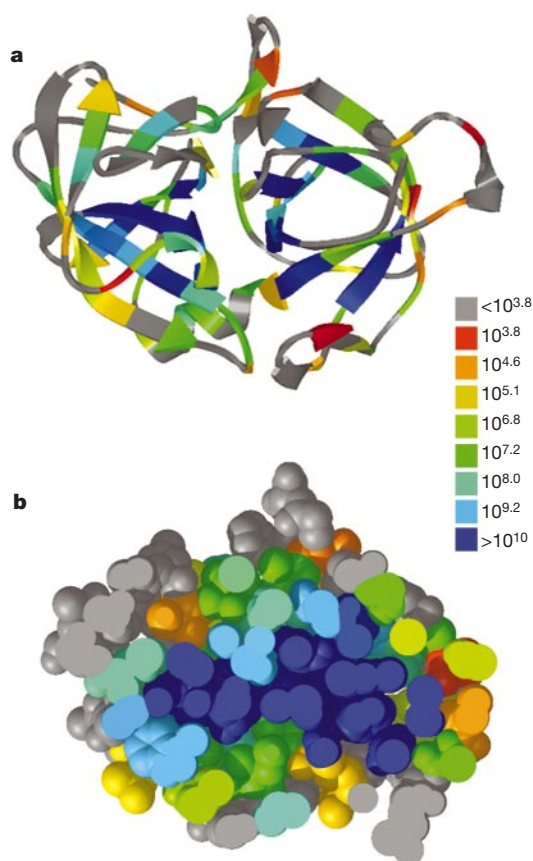


Figure 2 Hydrogen-exchange protection of α LP. **a**, **b**, Ribbon (**a**) and space-filling (**b**) representations of α LP coloured by amide protection factors. Residues with amide protection factors $<10^{3.8}$ are coloured grey. Residues with protection factors of $10^{3.8}$ to $\geq 10^{10.7}$ are coloured from red to blue. (A table listing the calculated protection factors at each pH is included as Supplementary Information.) The entire protease molecule is shown in **a**. A slice through the middle of the molecule in **b** reveals protection patterns in its interior. Figures were generated using MIDAS²⁴.

Table 1 Survival assay rates

pH	Protease	Inactivation rate (day ⁻¹)	Cleavage rate (day ⁻¹)
5	αLP	0.005 ± 0.001	0.01 ± 0.002
5	Trypsin	0.5 ± 0.02	0.29 ± 0.05
5	Chymotrypsin	0.06 ± 0.01*	0.072 ± 0.05*
7	αLP	0.006 ± 0.001	0.009 ± 0.007
7	Trypsin	0.83 ± 0.05	0.8 ± 0.2
7	Chymotrypsin	1.43 ± 0.08	0.9 ± 0.1
8	αLP	0.003 ± 0.001	ND
8	Trypsin	0.62 ± 0.05	ND
8	Chymotrypsin	2.22 ± 0.07	ND

Inactivation values indicate the rate of loss of enzyme activity on chromogenic substrate. Cleavage values indicate the rate of loss of a band on gel. The pH independence of the rate of αLP inactivation indicates that the unimolecular unfolding reaction is rate-limiting (see text), not the bimolecular cleavage reaction. ND, not determined.

* Chymotrypsin dimerizes at pH 5 (ref. 23), leading to an anomalously slow rate of inactivation (see data at pH 7 and 8 for typical behaviour).

homologues lacking pro-regions)¹. High glycine content would increase the entropy of unfolded states, but could also lead to greater native-state rigidity and increased proteolytic resistance by allowing tighter turns and tighter packing.

To assess whether the unusual dynamic properties of αLP sufficiently restrict conformational fluctuations to enhance proteolytic stability, we directly compared the lifetimes of αLP and its thermodynamically stabilized^{11,12} mammalian homologues, chymotrypsin and trypsin, under highly proteolytic conditions. Survival assays were set up in which equimolar amounts of all three proteases were mixed and incubated at pH 5, 7 and 8, providing a nearly 1,000-fold variation in proteolytic activity. Because the substrate specificities of the enzymes are orthogonal, it was possible to assay the survival of each protease over time by measuring the enzymatic activity of the mixture on three different chromogenic substrates, specific for each protease. The biological activity of αLP remains virtually unchanged, whereas its mammalian counterparts are readily destroyed at rates ≥100-fold faster than αLP (Fig. 3a and Table 1). This is not due to a greater number of, or increased accessibility to, proteolytic sites on chymotrypsin and trypsin. In fact, when normalized to the number of residues, αLP has the same percentage of sites (defined as ≥20% surface accessibility) as

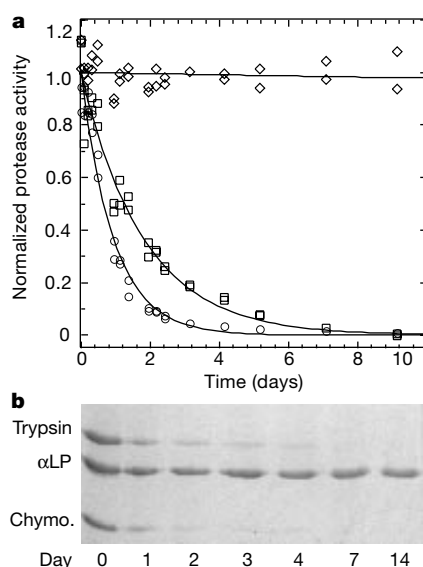


Figure 3 Survival assay at pH 7, 25 °C. **a**, αLP (diamonds) is more resistant to proteolysis than chymotrypsin (circles) and trypsin (squares), as monitored by loss of proteolytic activity of the mixture on three different chromogenic substrates specific for each protease. **b**, Quantification of the loss of intact protein monitored by SDS–polyacrylamide gel electrophoresis.

chymotrypsin and ~30% more than trypsin. Nor is the longer lifetime of αLP due to an ability to remain active when cleaved. At pH 7, the rates of cleavage monitored by gel electrophoresis for all three proteases are within twofold of their inactivation rates (Fig. 3b and Table 1). Therefore, αLP has the ability to withstand proteolysis and autoproteolysis for months longer than its thermodynamically stabilized counterparts.

We next determined whether αLP has a protease resistance that is optimized to the greatest extent possible. Given the very slow rate of αLP global unfolding¹, the most sensitive measure of optimization for proteolytic resistance is whether local unfolding transitions that could lead to proteolysis at a faster rate than through global unfolding have also been suppressed. This was examined quantitatively by comparing the rate of inactivation of αLP through autoproteolysis, measured by loss of enzymatic activity, to its rate of global unfolding in the absence of proteolysis, measured by following the fluorescence decrease during unfolding of an inactive variant, S195A αLP (active site Ser replaced by Ala) (Fig. 4). The rate of αLP autoproteolysis is 0.04 day⁻¹; the global unfolding rate is 0.02 day⁻¹. The twofold difference in rate may be due to a greater stability of the variant, but in any case corresponds to a free-energy difference of only 0.3 kcal mol⁻¹. This indicates that the extraordinary rigidity at the molecular level demonstrated by H–X successfully restricts local unfolding fluctuations so that only global unfolding, which occurs on a year timescale, leads to proteolytic accessibility.

These results provide the first functional rationale, to our knowledge, for the existence of pro-region-dependent folding. The observed suppression of both local and global unfolding transitions must be the direct consequence of an almost perfectly cooperative barrier to unfolding. Such extreme cooperativity comes at a significant energetic cost: the native state of αLP is not thermodynamically stable and can be reached only with the help of its pro-region. Thus, the development of desirable native-state properties has been made possible only by the transient existence of the pro-region, which has allowed the energetic landscape of the native state to evolve independently of the folding landscape.

Kinetic stability provides a novel mechanism for the evolution of optimal functional properties. When coupled with pro-region-dependent folding it provides for enhanced longevity in degradative environments not only for αLP, but probably for a wide array of pro-region-containing proteases and other secreted enzymes. In addition, kinetic stability can function as a timer to regulate biological function in non-proteolytic proteins that have an efficient folding pathway¹³. Examples include serpins¹⁴, membrane fusion proteins such as influenza haemagglutinin^{15,16} and possibly HIV gp41 (ref. 17), and heat-shock transcription factor¹⁸. Kinetic stability is also a fundamental aspect of many human amyloid diseases, in

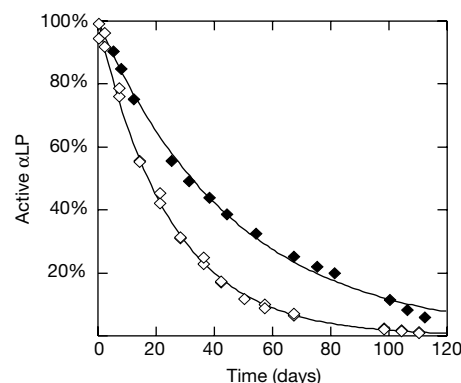


Figure 4 Autolysis compared with global unfolding rate. At 25 °C, 1 M GdnHCl, where local unfolding is promoted, the rate of autoproteolysis of the active enzyme (open symbols) monitored by loss of proteolytic activity is less than twofold faster than the rate of global unfolding of the inactive enzyme (filled symbols) monitored by loss of fluorescence.

which a barrier separating the native conformation from an inactive or even toxic conformation is surmounted over time or lowered through mutation or perturbed conditions¹⁹. The use of kinetic stability highlights the importance of the role of protein-folding energetics and dynamics, beyond simply specifying the active conformation, in critically influencing biological activity and lifetime. □

Methods

Protein expression and purification

α LP and S195A α LP were expressed in *Escherichia coli* and purified as described^{1,20}. Rat trypsinogen was expressed in *Pichia* as described²¹. Secreted protein auto-activated to trypsin, which was purified as follows. All buffers contained 10 mM calcium chloride. The supernatant from cells was slowly brought to 4.5 M NaCl, loaded onto a phenyl sepharose fast-flow column, washed with 50 mM MES at pH 6.0, 4.5 M NaCl then eluted with 50 mM MES at pH 6.0. After dialysing to remove NaCl, the eluted protein was loaded onto a p-amino benzamidine agarose column, washed with 50 mM Tris at pH 7.5 and eluted with 100 mM acetic acid into fractions containing 1/20 volume 0.5 M sodium citrate at pH 4.5. Fractions with active protein were pooled and dialysed into 1 mM HCl and assayed for activity as described²¹.

Hydrogen exchange

Amide exchange of wild-type α LP was initiated by the addition of protonated 15N-labelled protease to one of the following buffers containing >97% D₂O: 100 mM NaPO₄, 100 mM NaCl and 2.5 mM NaOAc(d₃) at pH 6.84; 100 mM Tris (deuterated), 100 mM NaCl and 2.5 mM NaOAc(d₃) at pH 8.04; or 100 mM NaBO₃, 100 mM NaCl and 2.5 mM NaOAc(d₃) at pH 9.17. Exchange occurred at room temperature and was monitored over a period of 6 months by two-dimensional HSQC (heteronuclear single-quantum correlation spectroscopy) experiments using a Varian 600 MHz NMR spectrometer. Fit peak heights were taken from the processed data and used to calculate the observed exchange rates for each amide proton using a single decaying exponential function. The observed rates were converted into protection factors for each amide using calculated intrinsic exchange rates²². Amide assignments at high pH were confirmed by comparing the NOESY (nuclear Overhauser enhancement spectroscopy) crosspeaks with those from a lower pH spectrum. The minimum observable exchange rate was 0.004 day⁻¹. Amides with exchange rates measured at several pH values demonstrated expected EX2 pH dependencies⁶, with the exception of five amides, four of which cluster near the active site. A full table of exchange rates as a function of pH is included as Supplementary Information.

Survival assays

A concentration of 6.5 μ M α LP, trypsin and chymotrypsin (TLCK (tosyl-L-lysine chloromethyl ketone)-treated; Worthington) were incubated together at 25 °C in 10 mM CaCl₂ and 50 mM KOAc pH 5.0, 50 mM MOPS at pH 7, 50 mM HEPES at pH 7, or 50 mM Tris at pH 8. Aliquots were removed over time and the survival of the individual proteases was measured on the basis of their activities, which could be distinguished given their non-overlapping specificities for different substrates (succinyl-Ala-Pro-Ala-pNA, succinyl-Ala-Ala-Pro-Arg-pNA, succinyl-Ala-Ala-Pro-Leu-pNA, used for α LP, trypsin and chymotrypsin, respectively, all substrates at 1 mM in 10 mM CaCl₂, 100 mM Tris at pH 8). Cleavage of the proteases during survival assays was monitored by quantifying the disappearance of the full-length species as monitored by band intensity in an SDS-polyacrylamide gel stained with Coomassie, using ImageQuant for Macintosh v1.2.

The inactivation rate (k) in three separate experiments at each pH, and cleavage in single experiments at pH 5 and pH 7 were fit for each protease with a single exponential equation $y = a - b \exp(-kt)$. To normalize for slight concentration variations in separate survival assays, the raw data were treated by subtracting the offset (a) and dividing by the amplitude of the decrease (b). The normalized data from three separate survival assays for each protease at each pH were combined and fit with the single exponential equation to determine the inactivation rate and standard error given in Table 1.

Autolysis assay

α LP (3.25 μ M) was incubated in 1 M guanidinium (Gdn)HCl, 10 mM KOAc, pH 5.0, at 25 °C. α LP activity was followed by monitoring the level of hydrolysis of 1 mM succinyl-Ala-Pro-Ala-pNA in 100 mM Tris at pH 8.0 by aliquots removed over time.

Global unfolding

Fluorescence measurements of the inactive mutant S195A α LP at 1.75 μ M in 0.98 M GdnHCl, 10 mM KOAc at pH 5.0 were made with excitation at 283 nm, emission 322 nm in an 8100SLM-Aminco fluorimeter connected to an external waterbath set to 25 °C. The sealed fluorescence cuvette was maintained between measurements in a 25 °C incubator.

Received 2 August; accepted 6 November 2001.

- Sohl, J. L., Jaswal, S. S. & Agard, D. A. Unfolded conformations of α -lytic protease are more stable than its native state. *Nature* **395**, 817–819 (1998).
- Baker, D., Shiau, A. K. & Agard, D. A. The role of pro regions in protein folding. *Curr. Opin. Cell Biol.* **5**, 966–970 (1993).
- Perona, J. J. & Craik, C. S. Structural basis of substrate specificity in the serine proteases. *Protein Sci.* **4**, 337–360 (1995).
- Fontana, A., Polverino de Laureto, P., De Filippis, V., Scaramella, E. & Zamboni, M. Probing the partly folded states of proteins by limited proteolysis. *Fold Des.* **2**, R17–R26 (1997).
- Chamberlain, A. K. & Marqusee, S. Touring the landscapes: partially folded proteins examined by hydrogen exchange. *Structure* **5**, 859–863 (1997).
- Englander, S. W., Sosnick, T. R., Englander, J. J. & Mayne, L. Mechanisms and uses of hydrogen exchange. *Curr. Opin. Struct. Biol.* **6**, 18–23 (1996).
- Rupley, J. A. Susceptibility to attack by proteolytic enzymes. *Methods Enzymol.* **11**, 905–917 (1967).
- Wang, L. & Kallenbach, N. R. Proteolysis as a measure of the free energy difference between cytochrome c and its derivatives. *Protein Sci.* **7**, 2460–2464 (1998).
- Li, R. & Woodward, C. The hydrogen exchange core and protein folding. *Protein Sci.* **8**, 1571–1590 (1999).
- Huyghe-Despointes, B. M., Scholtz, J. M. & Pace, C. N. Protein conformational stabilities can be determined from hydrogen exchange rates. *Nature Struct. Biol.* **6**, 910–912 (1999).
- Santorio, M. M. & Bolen, D. W. Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl α -chymotrypsin using different denaturants. *Biochemistry* **27**, 8063–8068 (1988).
- Wang, E. C., Hung, S. H., Cahoon, M. & Hedstrom, L. The role of the Cys191–Cys220 disulfide bond in trypsin: new targets for engineering substrate specificity. *Protein Eng.* **10**, 405–411 (1997).
- Lee, C., Park, S. H., Lee, M. Y. & Yu, M. H. Regulation of protein function by native metastability. *Proc. Natl Acad. Sci. USA* **97**, 7727–7731 (2000).
- Huber, R. & Carrell, R. W. Implications of the three-dimensional structure of alpha 1-antitrypsin for structure and function of serpins. *Biochemistry* **28**, 8951–8966 (1989).
- Carr, C. M. & Kim, P. S. A spring-loaded mechanism for the conformational change of influenza hemagglutinin. *Cell* **73**, 823–832 (1993).
- Bullough, P. A., Hughson, F. M., Skehel, J. J. & Wiley, D. C. Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature* **371**, 37–43 (1994).
- Chan, D. C., Fass, D., Berger, J. M. & Kim, P. S. Core structure of gp41 from the HIV envelope glycoprotein. *Cell* **89**, 263–273 (1997).
- Orosz, A., Wisniewski, J. & Wu, C. Regulation of *Drosophila* heat shock factor trimerization: global sequence requirements and independence of nuclear localization. *Mol. Cell Biol.* **16**, 7018–7030 (1996).
- Kelly, J. W. Alternative conformations of amyloidogenic proteins govern their behavior. *Curr. Opin. Struct. Biol.* **6**, 11–17 (1996).
- Mace, J. E. & Agard, D. A. Kinetic and structural characterization of mutations of glycine 216 in α -lytic protease: a new target for engineering substrate specificity. *J. Mol. Biol.* **254**, 720–736 (1995).
- Halfon, S. & Craik, C. S. Regulation of proteolytic activity by engineered tridentate metal binding loops. *J. Am. Chem. Soc.* **118**, 1227–1228 (1996).
- Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. Primary structure effects on peptide group hydrogen exchange. *Proteins* **17**, 75–86 (1993).
- Bolen, D. W. & Santoro, M. M. Unfolding free energy changes determined by the linear extrapolation method. 2. Incorporation of delta G degrees N-U values in a thermodynamic cycle. *Biochemistry* **27**, 8069–8074 (1988).
- Midas. The MIDAS display system. *J. Mol. Graph.* **6**, 13–27 (1988).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank Y. Shibata for assistance with hydrogen exchange experiments, and J. Harris, T. Baird and C. Craik for assistance with trypsin purification. S.S.J. was supported by a Howard Hughes Medical Institute predoctoral fellowship. Research funding was provided by Howard Hughes Medical Institute.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.A.A. (e-mail: agard@msg.ucsf.edu).

Automation information

Robotic systems and accessories for high-throughput applications.

Quick-Combi Sealer

HJ-Bioanalytik GmbH www.riplate.de

Well covered

This component of the Riplate microplate range can be used independently or as part of a robotic system to automatically seal all types of microtitre plates with aluminium foil or a 96-well lid in just 4 seconds. The unit is loaded with plate lids or a roll of aluminium foil. The plate then moves through the unit, the lid drops or the foil unwinds and cuts, and pressure is applied to produce the seal with no effect on the sample. The foil-cutting unit is available in manual and automatic versions. Both the foil and the adhesive are 100% DMSO resistance approved, and the foil is easily removed from the plate, leaving no trace. The Quick-Combi Sealer is suitable for use with Riplate deep-well plates, microtitre plates and 96 micro racks.



Foiled again: seal microplates with Quick-Combi.

and balances. The device is also suitable for quality control and process monitoring in the food, pharmaceutical, paper and petrochemical industries.

Automation Planner

Zymark www.zymark.com

Zap your apps

The Zymark Automation Planner (ZAP) is an interactive software program that allows users to browse its selection of automation setup scenarios and assemble the configuration that best fits their applications, throughput, capacity and time constraints. The level of automation can range from a stand-alone workstation to a fully integrated system. ZAP configures more than 600 automation solutions for over 200 applications clustered within 30 groups. The program also generates a virtual automation environment that is directly linked to Zymark's CLARA system integration software, allowing users to model their methods. Once the configuration is assembled, the same software can execute the methods to provide a real-time running solution. Specific application areas include compound logistics, genomics, proteomics, screening, lead optimization and ADMET.

Genesis Freedom

Tecan www.tecan.com

Out on a limb

Tecan's robotic laboratory workstation is a modular, scalable, high-throughput system for research in genomics, proteomics and drug discovery. The compact unit can accept up to three robotic arms and takes a wide selection of robotic manipulators and accessories, which can be combined to create up to nine different arm configurations. The arms can reach areas beyond the desktop, including underneath the workstation itself. Genomics applications include nucleic acid extraction and sample preparation, PCR setup, DNA normalization, genotyping, SNP scoring and automated *in situ* hybridization



High throughput from Tecan.

for high-throughput gene expression and functional studies. In proteomics research, the Genesis Freedom can be used to automate detection of unknown or low-abundant proteins, in-gel digestion, MALDI spotting and two-dimensional gel electrophoresis. It is also suitable for a wide spectrum of drug discovery applications, including early-stage compound synthesis, assay development, HTS and hit-picking, ADME and toxicology assays.

2100 Bioanalyzer A.02.01

Agilent Technologies www.agilent.com

LabChip system upgrade

A new calibration feature is included in revision A.02.01 of the Agilent 2100 Bioanalyzer. This allows users of the Protein 200 LabChip kit to perform absolute protein quantification based on external standard calibration. After the analysis of user-defined protein standards at different concentrations, the program automatically generates a linear regression curve that is applied to determine the absolute concentration of samples within the same chip. This feature allows protein sizing and absolute quantification to be carried out within a single experiment. The software for RNA analyses has also been improved in the recent revision, with alignment functionality now included. A lower marker included in the RNA 6000 Nano LabChip kit or one of the ribosomal bands can be used for sample alignment, enabling researchers to distinguish different types of RNA based on electrophoretic traces.

These notes are compiled in the Nature office from information provided by the manufacturers.

SM-IP21

Thermo Labsystems

www.thermolabsystems.com

That figures

This enhanced version of the SM-IP21 sample management software program enables users to configure integration logic with fully reusable components, thereby reducing the time needed for system setup and configuration. The integration server handles all transaction requests, executes the integration rules and ensures successful transfer of data. Configuration is achieved through a dedicated graphical user interface that provides a centralized environment in which integration rules are built, maintained and managed. Product installation is eased through an install shield with consequent automation of procedures that were previously manual.

DH Sample Changer 748

Metrohm www.metrohm.co.uk

Win by a head

Featuring a titration head that can move in three directions, Metrohm's PC-controlled sample changer is well-suited to use in water treatment and environmental analysis laboratories, where conductivity and pH measurements, as well as titration, have to be carried out on a series of samples. Two sample racks, each holding 24 beakers, occupy an area of just 70 × 50 cm. The sample volume in the beakers can vary between 10 ml and 250 ml. The system is controlled with TiNet 2.3 software running under Windows, used together with the company's Titrino titrators and other peripheral devices such as printers

Contacts

Publisher: Fabien Savenay
Editor: Paul Smaglik
Sales Director: Ben Crowe

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Group European Manager:

Leonie Welss (4954)

European Manager:

Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953), Ben Corp (4974),
Andy Douglas (4975)

Holland/ Italy:

Nevin Bayoumi (4978)

Scandinavia:

Sille Opstrup (4994)

Spain/ Portugal:

Leonie Welss (4954)

Production Manager:

Billie Franklin

To send films and materials use

London address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Laura Pearson (4977)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Offices

France/ Belgium:

Christine Niox-Chateau

Tel + 33 (0) 1 43 20 16 51

Fax + 33 (0) 1 43 20 51 52

e-mail: c.nioxchateau@nature.com

Germany/ Austria/ Switzerland:

Patrick Phelan/ Kate Turner

Tel + 49 89 54 90 57 11/2

Fax + 49 89 54 90 57 20

e-mails: p.phelan@nature.com

k.turner@nature.com

US Head Office, New York

345 Park Avenue South,

10th Floor, New York, NY 10010-1707

Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Director:

Ben Crowe

US Sales Manager:

Peyton Mason

US Advertising Coordinator:

Ashly de Leon

Japan Head Office, Tokyo

MG Ichigaya Building (5F),

19-1 Harakatomachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

e-mail: k.cowan@naturejp.com

Japan Manager:

Kate Cowan

naturejobs

Space to manoeuvre

Space. In the past months, I've heard scientists from several disciplines in multiple sectors from the United States, the European Union and Britain use this simple word to convey different things. None of the concepts involved could be described as exactly straightforward, even though the brevity of the word seems a proxy for certitude.

For example, structural biologists use it when discussing 'fold space', which could pertain to the number of shapes a particular protein can take on, the number of protein families, or the total number of proteins in the Universe (a different kind of space entirely). But industrialists use space to describe their strategic niche compared with those of their competitors.

Confusion grows when considering the word's usage in a structural genomics company. In these instances, it can be hard to tell if the chief executive is talking about the company's scientific strategy or its business plan.

I'm even less sure of the word's origins than I am of its precise usage in any scientific or business context. I suspect it started in mathematics or theoretical physics, then later crossed over into biology. This points to the physicist as the likely semantic vector. The theory makes some sense, as physicists dominated structural biology in its early days and also moved into the financial world, where the term's frequency is perhaps greatest.

Whatever the origins, the word's trajectory serves as both a linguistic artefact and a signpost. An artefact because its multiple uses suggest that science is becoming increasingly interdisciplinary. And a signpost because the varied meanings could serve as a warning about how difficult it can be for scientists in different disciplines and sectors to speak the same language.

Paul Smaglik
Naturejobs editor



Contents

SPECIAL REPORT

Toxicogenomics comes of age p4

WWW.NATUREJOBS.COM

Career centre
Information on the scientific
job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS



Will toxicogenomics turn toxicology into a predictive and preventive science? Diane Gershon looks at the plans for the future.

Why does one person contract an environment-related disease when their neighbour never shows as much as a symptom? This is one of several perplexing questions that a new national toxicogenomics research consortium hopes to answer.

The consortium was formed last November when the US National Institute of Environmental Health Sciences (NIEHS) in Research Triangle Park, North Carolina, awarded five-year grants totalling more than US\$37 million to five US academic research centres. Although the centres were selected on a competitive basis, the consortium offers participants the chance to work both collaboratively and independently.

Each will pursue a limited number of independent, investigator-initiated research projects focused on their own area of expertise. Additionally, the toxicology research cores at each member site will do 'directed' experiments that address major questions agreed upon by a steering committee. They will be assisted by NIEHS intramural scientists and by an outside research contractor.

These 'common' experiments will be done on a standardized platform of microarrays and the data will be deposited in a public database. The research contractor, who has yet to be appointed, will bring outside expertise in analysing RNA profiles and in database development.

ON THE JOB FRONT

Although toxicology draws from fairly diverse disciplines, the consortium will provide a unique opportunity for card-carrying toxicologists to rub shoulders with molecular biologists who have expertise in microarray gene-expression profiling and proteomic technologies. It will also require people who have the ability to shuttle back and forth between molecular biology and bioinformatics.

In recent years, there has been a push in toxicology towards molecular biology. But the consortium creates a 'new nexus' for these groups to work together, says Helmut Zarbl of the Fred Hutchinson Cancer Research Center in Seattle, Washington. This institute, partnered with the University of Washington, is one of the five

participating centres in the new consortium. Zarbl's research interests centre on chemical carcinogenesis and differential susceptibility — he is currently using a rat model to study environmental susceptibility to breast cancer.

Some recruiting for the centres is expected over the five-year period. But, when split five ways, the grant amount received by individual sites does not allow room for many new appointments, as the microarray technology that will be used for much of the work is still quite expensive.

Peter Spencer, who is leading the consortium team at the Oregon Health and Science University in Portland, expects to recruit a couple of staff over the next two years specifically for the consortium effort. He has plans for a new microarray core facility and will be looking for a director to run it.

MULTIFACETED ATTRACTION

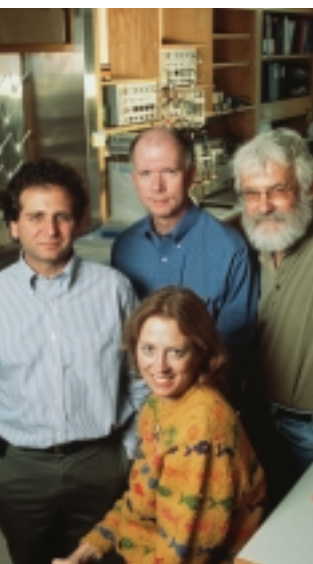
The chance to take part in a multifaceted approach to disease development appealed to David Schwartz, head of pulmonary medicine at Duke University in Durham, North Carolina, and principal investigator for the Duke effort. The tendency until now, he says, has been to focus on either environmental exposures or genetic causes.

The consortium allows for a more complex view of disease aetiology. "We are all blinded to a certain extent by our own niche," he says. Schwartz's interest in getting beyond his own niche attracted him to Duke in the first place, because of its commitment to developing the area of genomics in general, through the newly established Institute for Genomics Sciences and Policy, and specifically environmental genetics, in which at least five additional recruitments are expected.

Having five centres working together, while also pursuing their own lines of inquiry, allows each to play to its existing strengths. For example, the University of North Carolina at Chapel Hill site will concentrate on susceptibility factors in the genomic response to toxicants, with a special emphasis on environmental carcinogenesis.

"We'd like to identify every gene that responds 'up' or 'down' in response to oxidative stress in at least three different kinds of human cells — epithelial, fibroblasts and lymphoblasts," says William

David Schwartz (left) relishes the multifaceted approach that the new toxicogenomics initiative offers.





Driven by technology

Kaufmann, a professor of pathology and laboratory medicine. "And, moreover, to relate these patterns of response to biological responses monitored at the level of DNA damage/repair and cell-cycle checkpoints."

The Massachusetts Institute of Technology, the fifth participating centre, will use gene-expression profiling to study the effects of environmental alkylating agents on human health.

TOXICOGENOMICS TIMELINE

Although each centre will have its own specialism, initial work will concentrate on setting standards for the technology that all the members will abide by, says Michael McClure, chief of the organs and systems toxicology branch in the NIEHS's division of extramural research and training. McClure anticipates that this phase-one 'harmonization' effort, which will establish clear definitions of standards and practices, precision of data and quality control, will take the better part of the first year to complete.

Phase two will focus on the development of a national microarray gene-expression database and the inclusion of data from a series of known toxicants in base-line species conducted in at least three of the five participating centres. Later phases will see an increase in the number of experiments, to include a series of unknowns, and an opening up of the database to data from investigators outside the consortium using some sort of clearing-house mechanism.

On the broader front, it is hoped that the initiative will lay some of the foundations for improved human risk assessment, as well as providing a clearer insight into the pathways and mechanisms of toxicity. It is also hoped that it will eventually lead to better exposure assessment and the development of biomarkers to detect exposures before symptoms arise.

By studying gene-expression patterns in individuals who are and are not susceptible, researchers hope to be able to elucidate the mechanisms of susceptibility to different environmental agents, which may in turn lead to improved treatments. Another potential spin-off of the research is that it could lead to a reduction in animal testing.

The consortium's ability to meet both short- and long-term goals will dictate whether it creates more

Toxicogenomics is a relatively new but rapidly growing research area that could revolutionize the field of toxicology. Central to this approach are microarray and proteomic technologies that, when coupled with robust bioinformatics analyses, provide a global picture of how organisms respond to environmental toxicants, be they chemicals, physical agents or physiological stresses.

This new line of research should provide a clearer understanding of the role of gene-environment interactions in disease, and should shed light on why some individuals are unaffected by particular environmental factors whereas others succumb to disease.

The completion of the draft sequence of the human genome and the rapid progress in other genome-sequencing

projects, coupled with powerful technologies such as gene-expression profiling using microarrays, has helped to propel toxicogenomics forwards.

With unprecedented accuracy and speed, researchers can determine the relative levels of expression of thousands of genes at a time. A standard microarray, for example, now contains 18,000–20,000 genes. Microarray gene-expression profiling provides a specific pattern, or 'signature', of all the genes that are active or inactive in the presence of a particular toxicant. One of the specified aims of the toxicogenomics research consortium, formed last November by the National Institute of Environmental Health Sciences, is to build a public database of known toxicant signatures. **D.G.**

Searching for signatures: Helmut Zarbl (above) and Michael McClure (seated on the right).

research opportunities, both within the five member institutions and the NIEHS, and in establishing new collaborations beyond the existing framework.

"It's anticipated that there will be additional components added to this network," says McClure. ■
Diane Gershon is assistant editor, new technology at *Nature Medicine*.

Web links

Duke University ♦ www.duke.edu

Fred Hutchinson Cancer Research Center ♦ www.fhcrc.org

Massachusetts Institute of Technology ♦ www.mit.edu

Oregon Health and Science University ♦ www.ohsu.edu/croet

University of North Carolina, Chapel Hill ♦ www.unc.edu

NIEHS, Research Triangle Park ♦ www.niehs.nih.gov/nct/home.htm